



PHD

## Learning through graphical timelines

Martin, David

*Award date:*  
2018

*Awarding institution:*  
University of Bath

[Link to publication](#)

## Alternative formats

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

### Take down policy

If you consider content within Bath's Research Portal to be in breach of UK law, please contact: [openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk) with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Learning through graphical timelines

Volume 1 of 1

David John Martin

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Psychology

May 2018

# COPYRIGHT

Attention is drawn to the fact that copyright of this thesis/portfolio rests with the author and copyright of any previously published materials included may rest with third parties. A copy of this thesis/portfolio has been supplied on condition that anyone who consults it understands that they must not copy it or use material from it except as permitted by law or with the consent of the author or other copyright owners, as applicable.

This thesis/portfolio may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation with effect from.....(date)

Signed on behalf of the Faculty/School of Psychology.....

# Contents

<b>1</b>	<b>Introduction</b>	<b>17</b>
1.1	Aims, research questions and objectives of the research . . . . .	18
1.2	Structure of the thesis . . . . .	20
1.3	Summary . . . . .	23
<b>2</b>	<b>Literature Review</b>	<b>24</b>
2.1	Chapter overview . . . . .	24
2.2	Review of the literature . . . . .	25
2.3	Summary and Research Context . . . . .	52
2.4	Definitions . . . . .	54
2.5	Aims and Overall Hypotheses . . . . .	55
2.6	Relationship between Studies and Hypotheses . . . . .	56
2.7	Summary . . . . .	56
<b>3</b>	<b>Study One: A study of using a timeline in a UK secondary school</b>	<b>60</b>
3.1	Chapter Overview . . . . .	60
3.2	Introduction . . . . .	61
3.2.1	Rationale . . . . .	61
3.2.2	Aims . . . . .	66
3.3	Method . . . . .	66
3.3.1	Participants and Recruitment . . . . .	66

3.3.2	Materials and Measures . . . . .	67
3.3.3	Design . . . . .	70
3.3.4	Procedure . . . . .	71
3.4	Results . . . . .	74
3.4.1	Year seven results . . . . .	74
3.4.2	Year eight results . . . . .	74
3.4.2.1	Group comparisons of year eight data . . . . .	74
3.4.2.2	Analysis of year eight results . . . . .	74
3.4.3	Year nine results . . . . .	75
3.4.3.1	Group comparisons of year nine data . . . . .	75
3.4.3.2	Analysis of year nine results . . . . .	75
3.5	Discussion . . . . .	76
<b>4</b>	<b>Study Two: An experimental study on the use of timelines</b>	<b>84</b>
4.1	Chapter Overview . . . . .	84
4.2	Introduction . . . . .	85
4.2.1	Rationale . . . . .	85
4.2.2	Spatial and Visualisation Abilities . . . . .	89
4.2.3	Aims . . . . .	91
4.3	Method . . . . .	92
4.3.1	Participants and Recruitment . . . . .	92
4.3.2	Materials and Measures . . . . .	93
4.3.2.1	Experiment framework . . . . .	93
4.3.2.2	Overall Experiment structure . . . . .	94
4.3.2.3	Text and timeline materials . . . . .	94
4.3.2.4	Comprehension and immediate retention questions . . . . .	96
4.3.2.5	Self reported effort and difficulty . . . . .	96
4.3.2.6	Plymouth Sensory Imagery Questionnaire (PSI-Q) . . . . .	97

4.3.2.7	Learning Scenario Questionnaire (LSQ)	97
4.3.2.8	Multimedia Learning Preference Questionnaire (MLPQ)	97
4.3.2.9	Mental rotation ability	97
4.3.2.10	Analytical approach	97
4.3.3	Design	99
4.3.4	Procedure	100
4.4	Results	101
4.4.1	Group balance	101
4.4.2	Participant Accuracy	101
4.4.3	Participant Response Times	102
4.4.4	Other Variables	103
4.4.5	Comparison of the two approaches to cognitive load measurement	104
4.4.6	Overall Timing	105
4.5	Discussion	106

<b>5</b>	<b>Study Three: An experimental study on the use of timelines with delayed assessment</b>	<b>110</b>
5.1	Chapter Overview	110
5.2	Introduction	111
5.2.1	Rationale	111
5.2.1.1	Delayed retention test	112
5.2.1.2	Broader assessment of spatial ability	113
5.2.1.3	Increased assessment of mental rotation	113
5.2.2	Aims	114
5.3	Method	115
5.3.1	Participants and Recruitment	115
5.3.2	Materials and Measures	116
5.3.2.1	Experiment framework	116

5.3.2.2	Overall Experiment structure . . . . .	116
5.3.2.3	Text and timeline materials . . . . .	117
5.3.2.4	Mental rotation ability . . . . .	117
5.3.2.5	Object location memory test . . . . .	120
5.3.2.6	Demographic and Learning Preference questions . . . . .	120
5.3.2.7	Analysis approach . . . . .	120
5.3.3	Design . . . . .	121
5.3.4	Procedure . . . . .	122
5.4	Results . . . . .	124
5.4.1	Group balance . . . . .	124
5.4.2	Participant Accuracy . . . . .	124
5.4.3	Participant Response Times . . . . .	125
5.4.4	Other Variables . . . . .	127
5.4.5	Overall Timing . . . . .	128
5.5	Discussion . . . . .	128

## **6 Study Four: An experimental study examining text versus timelines and self**

	<b>creation of each type</b>	<b>132</b>
6.1	Chapter Overview . . . . .	132
6.2	Introduction . . . . .	134
6.2.1	Rationale . . . . .	134
6.2.2	Aims . . . . .	136
6.3	Method . . . . .	136
6.3.1	Participants and Recruitment . . . . .	136
6.3.2	Materials and Measures . . . . .	137
6.3.2.1	Materials . . . . .	137
6.3.2.2	Analysis of text used . . . . .	138
6.3.2.3	Additional Summary . . . . .	138

6.3.2.4	Additional Timeline . . . . .	139
6.3.2.5	Equipment for participant creation of additional material .	139
6.3.2.6	Computer test questions . . . . .	139
6.3.3	Design . . . . .	140
6.3.4	Procedure . . . . .	141
6.4	Results . . . . .	142
6.5	Discussion . . . . .	144
<b>7</b>	<b>Overall Discussion</b>	<b>147</b>
7.1	Chapter overview . . . . .	147
7.2	Introduction . . . . .	147
7.3	Summary of the research . . . . .	151
7.4	Reflecting on the hypotheses . . . . .	159
7.5	To what extent can the CTML principles be applied to timelines? . . . . .	165
7.6	Should timelines be considered as maps of time or as diagrams? . . . . .	169
7.7	Can response times be used as an indicator for cognitive load? . . . . .	171
7.8	Is there a relationship between timeline benefits and spatial abilities, or learning preferences? . . . . .	172
7.9	Reflections, implications, and future work . . . . .	172
7.10	Final comments . . . . .	176
	<b>Appendices</b>	<b>233</b>
<b>A</b>	<b>Key Principles of the Cognitive Theory of Multimedia Learning</b>	<b>234</b>
<b>B</b>	<b>Materials Used in Study One</b>	<b>237</b>
<b>C</b>	<b>Common Materials for Studies Two and Three</b>	<b>241</b>
<b>D</b>	<b>Additional Materials for Study Three</b>	<b>261</b>

<b>E</b>	<b>Materials for Study Four</b>	<b>265</b>
<b>F</b>	<b>Additional Tables for Study One</b>	<b>271</b>
<b>G</b>	<b>Additional Analysis for Study One</b>	<b>273</b>
<b>H</b>	<b>Additional Tables for Study Two</b>	<b>276</b>
<b>I</b>	<b>Additional Tables for Study Three</b>	<b>279</b>



# List of Figures

2.1	<i>Example of a timeline containing images . . . . .</i>	39
2.2	<i>Example of a simple timeline . . . . .</i>	59
3.1	<i>Example of a classroom timeline . . . . .</i>	64
3.2	<i>Example of a student produced timeline . . . . .</i>	65
4.1	<i>Study Two - Experiment structure . . . . .</i>	95
5.1	<i>Study Three - Experiment structure - first session . . . . .</i>	118
5.2	<i>Study Three - Experiment structure - second session, delayed retention . . . . .</i>	119
6.1	<i>Study four - interaction between mode and session . . . . .</i>	144
B.1	<i>Year 7 Religious changes in Tudor times . . . . .</i>	238
B.2	<i>Year 8 English Language Timeline . . . . .</i>	239
B.3	<i>Year 9 WW2 Timeline . . . . .</i>	240
C.1	<i>Brontavia timeline . . . . .</i>	245
C.2	<i>Spotania timeline . . . . .</i>	246
C.3	<i>Two digit number - graphics version . . . . .</i>	247
C.4	<i>Four digit number - text version . . . . .</i>	252
C.5	<i>Two item choice - graphics version . . . . .</i>	253
C.6	<i>Four item choice - text version . . . . .</i>	254

C.7	<i>Ordering - graphics version</i>	255
C.8	<i>Vividness - Plymouth Sensory Image Questionnaire</i>	256
C.9	<i>Example learning scenario question</i>	257
C.10	<i>Multimedia Preference Screen one</i>	257
C.11	<i>Multimedia Preference Screen two</i>	258
C.12	<i>Multimedia Preference Screen three</i>	259
C.13	<i>Mental rotation test example</i>	259
C.14	<i>Example retention question</i>	260
D.1	<i>Study Three: Letter rotation example</i>	261
D.2	<i>Study Three: Example screen for object location memory test (original in colour)</i>	264
D.3	<i>Study Three: Updated mental rotation instructions</i>	264
E.1	<i>Study Four: Summary timeline</i>	270

# List of Tables

2.1	Examples of Varied Uses of Timelines . . . . .	30
2.2	Hypotheses and their Associated Studies . . . . .	57
3.1	Quotations From the Primary Education Journal of the Historical Association Promoting Timeline Usage . . . . .	62
3.2	Study One - Numbers in Each Group . . . . .	67
3.3	Summary of Characteristics of Timelines used in Study One . . . . .	69
3.4	National Curriculum Sub Levels . . . . .	71
3.5	Conversion of Grades to Average Point Scores for Analysis . . . . .	72
4.1	Study Two: Demographics . . . . .	92
4.2	Study Two: Participant Highest Education Level . . . . .	92
4.3	Study Two Correlation Table . . . . .	104
4.4	Study Two: Effort and Response Time Correlations for Different Question Types . . . . .	105
5.1	Study Three - Group allocations . . . . .	116
5.2	Study Three Correlation Table . . . . .	127
6.1	Study four - Group summaries . . . . .	137
6.2	Study four - Group materials . . . . .	140
7.1	Summary of Characteristics of Timelines used in the studies . . . . .	158

7.2	Hypotheses and support from Studies . . . . .	164
A.1	Cognitive Theory of Multimedia Learning Principles Used in the Studies . . .	235
A.2	Cognitive Theory of Multimedia Learning Principles Used in the Studies - Continued . . . . .	236
C.1	Analysis of the Primary Text Used for Studies Two and Three . . . . .	244
C.2	Studies Two and Three: Comprehension Questions for Brontavia Story . . .	248
C.3	Studies Two and Three: Comprehension Questions for Spotania Story . . .	249
C.4	Studies Two and Three: Immediate Retention Questions . . . . .	250
C.5	Learning scenario Questions . . . . .	251
D.1	Study Three: Delayed Retention Questions for Brontavia Story . . . . .	262
D.2	Study Three: Delayed Retention Questions for Spotania Story . . . . .	263
E.1	Study Four questions . . . . .	269
F.1	Study One - Previous term marks . . . . .	271
F.2	Study One - Target mark distributions . . . . .	272
F.3	Study One - Current term mark distributions . . . . .	272
F.4	Study One - Calculated mark difference distributions . . . . .	272
H.1	Study Two - Variance, normality, skew, and kurtosis . . . . .	276
H.2	Study Two - Story comparison . . . . .	277
H.3	Study Two - Comprehension comparison . . . . .	277
H.4	Study Two - Retention comparison . . . . .	277
H.5	Study Two - Comprehension RT comparisons . . . . .	277
H.6	Study Two - Retention RT comparisons . . . . .	278
I.1	Study Three - Variance, normality, skew, and kurtosis . . . . .	279
I.2	Study Three - Comprehension RT comparisons . . . . .	280

I.3	Study Three - Initial retention RT comparisons . . . . .	280
-----	--	-----

# Acknowledgements

The work described in this thesis stemmed initially from an involvement in schools as a governor, and an interest in the technology used to support learning in schools, which then led to a desire to understand the psychology behind what makes learning technologies effective. I was lucky therefore to find, in the CREATE laboratory at the University of Bath, under the leadership of Professor Danaë Stanton Fraser, a group of researchers looking, from a wide range of perspectives, at these and other equally exciting technology topics. Being awarded a university studentship to help with the funding also made it possible to start on this journey of investigation. It has certainly been a long journey, as a part time student, with a full time job that involves a great deal of travel, I could not have completed this thesis without the flexibility, tolerance, and unswerving support that I have received. I know that almost everyone starts by giving credit to their supervisors, but for me Professor Danaë Stanton Fraser, and Professor Julie Turner-Cobb have definitely gone above and beyond what I could have expected, in their support and encouragement and, particularly in their patience!

The CREATE lab team have been a great help, it can be very lonely being a part time student, lab meetings, discussions and socials have been a lifeline.

I also thank the school teachers, school students, and all of the other participants in my studies.

My children have all supported me throughout the work. I thank Daniel for help with tricky programming problems, and Vikki and Fiona for proof reading my study materials. The

biggest thanks in this regard have to go to Dr Susanna Martin, who in a strange role reversal after the years of asking me for advice on school homework, ends up being the one in whose footsteps I am following with this thesis!

I, of course, save my greatest thanks to last, in order to thank my wife who has supported me throughout the whole process. Putting up with my working abroad for a great deal of the time and then seeing me disappearing to study for much of the rest, has not been easy I know, and having to postpone our 40<sup>th</sup> wedding anniversary while I wrote this thesis just added to that! I love you, and am eternally grateful for everything that you have given me.

# Abstract

This thesis investigates whether graphical timelines assist the process of learning. Promoted for the teaching of history, where they are considered to provide learners with a "Map of Time", and help them understand the "Big Picture", there is little research available concerning their benefits. Considering timelines as examples of multimedia, the thesis examines how the principles of multimedia learning, often directed more towards learning in science and technology, could be applied to timelines. The first study was performed in a UK secondary school. Setting the scene, with a focus on ecological validity, in examples of real world history teaching it compared the results from groups provided with additional timeline materials, and control groups without. Finding that timelines did increase comprehension in one year group, the second and third studies built upon this result, exploring effects on retention as well as comprehension. Under laboratory rather than classroom conditions, fictional materials were used to remove effects of prior knowledge. The experiments were extended by examining participant response times as a way of assessing the cognitive load associated with the material. The cognitive load measure was found to correlate with a common self report instrument, and also indicated a lower load for the timeline material. The results for retention accuracy varied between the studies but suggested that the expected multimedia effect was not present. Where these studies had allowed the participant to set the pace, the final study returned to a more school like condition, with a fixed study time, and explored self generation of timelines, however no multimedia effect was found. The thesis concludes with reflections on the work and



recommendations of future study ideas.

# Chapter 1

## Introduction

History, and its teaching, is subject to a great deal of political and societal pressure, mostly around the content but, also involving how it is presented. For many years schools have been encouraged to use timelines to provide a chronological overview and teach the “big picture”. In 2011 a government advisor commented that "much can be achieved now in improving pupils’ sense of time by teachers regularly using timelines" (Maddison, 2011, p. 9). While timelines are frequently employed by teachers as a teaching aid, and the government points to their use as beneficial, there is, however, very little evidence in the literature in support of their benefits as a teaching tool. Timelines are perhaps popular because they have an intuitive appeal as a connection between time and space, which the literature chapter discusses as being a deep rooted relationship (albeit with some aspects, such as the direction of flow, being affected by culture), but again there has been relatively little research into their effectiveness.

In order to understand whether timelines can help students to understand and learn history, it is also important to consider learning itself, and how it can be understood, measured and ultimately improved. There are a wide range of theories of learning, ranging from those of theorists such as Piaget and Vygotsky, who considered the practicalities of learning, and the importance of building on existing knowledge through scaffolding, with support from peers

or teachers, to theories of learning as memory, considering short and long term memory, how information is retrieved, and the cognitive load associated with learning new concepts. In this thesis the focus is on comprehension and retention of information, measuring firstly whether a learner can extract key information from a source (comprehension), and secondly whether this information is effectively retrievable at a later point (retention). It does not consider expert learning which would require the learner to transfer the material to existing concepts through scaffolding, enabling them to apply the knowledge to other concepts. Within the literature review, this thesis notes that cognitive load is often used as a measure of how effective a resource is for learning. Typically research has used self report tests, dual tasks, or biological measures to put a value on cognitive load. It has been speculated that response times during assessment could also give an indication of the relative cognitive load a student is experiencing.

Therefore this thesis has two key aims, the first being to establish whether timelines help learners understand historical information. The second aim was derived from the literature around learning and cognitive load; the key question here is whether question response times during comprehension of the different types of material can be used in place of self reported cognitive load.

## **1.1 Aims, research questions and objectives of the research**

The research described in this thesis used a quantitative approach to address the research questions. The majority of the studies were laboratory based but the opportunity was taken in study one to explore real-world usage of timelines in a school environment.

The primary aim of the studies was to compare the effects on learning of the presentation of learning material with a high temporal content in graphical form (as a timeline) and textual form (as a written story). A supplementary aim was to establish if response times could be used as an effective measure of the cognitive load that the material induces.

A review of the literature led to the following questions:

- Are timelines as effective as politicians and education advisors suggest, in particular is information presented in a timeline format more easily understood than text when people are learning?
- Furthermore, are timelines effective in promoting both comprehension and retention aspects of learning?
- Since timelines can be considered an example of multimedia combining both graphics and text, can a commonly used theory of multimedia learning - Mayer's Cognitive Theory of Multimedia Learning (CTML) be extended beyond its most frequent application in the fields of science, technology, engineering, and mathematics (STEM) to apply to timelines as well?
- When people are studying materials for learning, do their response times for comprehension questions correlate with self reported measures of cognitive load?

Five hypotheses were developed from these:

Hypothesis 1: Use of timelines will result in increased comprehension of information when compared to text.

Hypothesis 2: Use of timelines will result in shorter question response times during comprehension of information when compared to text.

Hypothesis 3: Use of timelines will result in greater retention of information when compared to text.

Hypothesis 4: Individual creation of timelines will result in greater retention of information when compared to provided timelines.

Hypothesis 5: Question response times will correlate positively with self reported estimates of cognitive load.

## **1.2 Structure of the thesis**

### **Chapter one: Introduction**

This chapter helps to introduce the reader to the key topics considered in the thesis, providing a brief overview of the rationale behind the thesis, the methodology employed and an overview of each of the chapters.

### **Chapter two: Literature review**

The literature review begins with an introduction to time, and our cognitive understanding of it. From here the discussion extends into an overview of timelines, and in particular their use in schools. To situate the research further, the literature review goes on to consider the process of learning, with particular emphasis on cognitive models of learning such as the dual coding theory and cognitive load theory. The discussion then focuses upon the Cognitive Theory of Multimedia Learning (CTML), which is considered to be of most direct relevance to the studies in this thesis. It continues with a critical evaluation of the gaps in this research, in particular with research mostly limited to science and technology based subject material, concluding that there are key gaps in the research that this thesis could fill. The literature review also provides a review of additional factors that may influence learning through timelines such as spatial ability and motivation. The chapter concludes with a summary highlighting the gaps that have been identified, and the related research questions which have been generated. The five hypotheses are presented at this point, and a table is provided showing how the four studies link to the hypotheses.

### **Chapter Three - Study One: A Study of Using a Timeline in a UK Secondary School**

The first study discussed in this thesis was designed to explore how timelines can be used in a UK school. Designed in collaboration with teachers, this study compared two equivalent sets of students as they undertook a history module. In the experimental set the teacher

taught the material with the addition of extra timeline resources. In the control set the students were taught with only the traditional materials. The teacher provided previous assessment scores, predicted scores and assessment scores for the experimental module. The students were unaware of the study until after it had taken place, and, since the researcher did not visit the class during the study, nor take part in the teaching or assessment of the students, this approach provided clear ecological validity for the study. The results indicated that students in year eight in the experimental condition, who were provided with supplementary timelines, did show an improvement in their scores in comparison to their predicted learning trajectory, while those who were in the control condition did not. Reflection on this study led to a number of questions which studies two, three and four would seek to answer.

#### **Chapter Four - Study Two: An experimental study on the use of timelines**

The second study was designed to take place in a laboratory environment, as this would allow for the experiment to take place under better controlled conditions. Using a within participant design, participants were taught about two fictional historical subjects, one by means of timelines and one using text. Participants were scored on their comprehension (with materials present) and on their retention (with materials absent). To act as a distraction, and to gain additional data regarding individual differences that some other researchers had shown to affect the boundaries of CTML principles, the participants also undertook a series of cognitive ability tests relating to aspects such as spatial ability, learning preferences, and multimedia preferences. To explore the aim relating to cognitive load measures, participants were asked, at key points during the tasks, to report how much effort they were expending to answer the questions that they were posed during the experiment. This was then compared to their recorded response times for those questions. It was shown that response time correlated with self reported cognitive load and could be used as a useful measure in future studies. The results of the experiment indicated that timelines

did not have an accuracy effect on comprehension but did reduce response times. During retention testing however, the accuracy results suggested that text materials were of greater benefit. Further investigation was required in studies three and four.

### **Chapter Five - Study Three: An experimental study on the use of timelines with delayed assessment**

Building on the previous work, the third study used the same materials and design as study two. The key difference was the addition of a delayed retention test, and of additional spatial ability tests. The results of study three again indicated no accuracy difference in the comprehension task but, as in study two, text proved to be beneficial in retention, although this time only in the delayed retention test. Using response times as a measure of load the results indicated again that, during comprehension, the cognitive load was lower for the timeline condition.

### **Chapter Six - Study Four: An experimental study examining text versus timelines and self creation of each type**

The final study chose to focus in on the unexpected retention results, examining only retention, but it also sought to establish whether the generation of materials such as timelines had an impact on student learning. In this study participants were required to read a piece of text about a fictional battle which was augmented with supplementary information in one of four conditions (Provided Timeline, Provided Notes, Generated Timeline, Generated Notes). The participant was given a set period of time to review the common text and to review or generate the additional materials. All materials were then taken away and an immediate retention task was given. One week later the participants returned and undertook the delayed retention task. The results again were unexpected, with those who received supplementary information performing better than those who generated it. In terms of timelines compared to text, an interaction effect was found with session, whereby

although both material types showed a decrease in scores between the two assessment periods, this occurred to a lesser extent in the timeline condition, suggesting that timeline information could lead to an increased persistence of knowledge.

## **Chapter Seven**

The final chapter discusses the four studies and situates them in the literature. Key reflections are made, firstly on each of the five hypotheses, and secondly on key principles from the CTML. The discussion continues with a reflection on the limitations of this work, and implications for the future.

### **1.3 Summary**

It can be seen from this overview that a wide variety of research areas have informed this body of work, and should consequently benefit from the studies described in this thesis. Although the results have not always been as anticipated, they have been meaningful and can be used as foundations for further explorations. At a time when the UK public is seen to be questioning experts, it is important to remember that assertions made by those in power should always have evidence to support them, this thesis is a step in understanding the relationship between timelines and learning which should eventually lead to a more sound basis for their use.



# **Chapter 2**

## **Literature Review**

### **2.1 Chapter overview**

This chapter places the studies described in this thesis into a broader context of existing research. It begins with a description of the fundamental nature of time, observing that unlike the three physical dimensions, where senses of touch and vision can provide cognitive references, there appears to be no direct sensory mechanism in human (or animal) physiology concerning the passage of time. Links between time and spatial cognition, both in terms of language, and experimental findings, are discussed, leading to the notion that graphical timelines may, by linking to underlying cognitive models in different ways than equivalent textual representations, assist the comprehension and retention of temporal information. The promotion of use of timelines in teaching in UK schools is described, highlighting a lack of evidence of effectiveness to match their popularity. A discussion of the learning process, with a description of generally accepted underlying cognitive models, and their reflection of known limits to rates of learning, leads to a description of a set of key principles developed over many years of work relating to multimedia learning such as that involved when a graphical timeline is used in conjunction with text (Mayer, 2014a). Although these principles would seem to be of direct relevance to timelines as examples of

multimedia where text is combined with a diagrammatic structure, research in connection with the principles has, however, focused upon examples from science and technology, where the graphical representation (whether static or dynamic) is usually either a concrete image, or a diagrammatic illustration of a concept. Research regarding the particular case of combinations of geographical maps and text are also discussed but, again, the focus has been found to mostly concern maps with features that directly relate to tangible experience (landmarks, topology, etc.). Timelines which can be considered to be both "maps of time", and multimedia in nature, have rarely been studied. The gap between expectations concerning the use of timelines in education and the sparsity of supporting evidence is discussed, and the chapter reviews known studies that have examined the use of timelines in the field, concluding with identification of areas for study.

## **2.2 Review of the literature**

### **The link between time and space in the mind**

The importance of time can be demonstrated through analysis of spoken/written language. In English, for example, the word "time" is the most frequently used noun (Boroditsky, 2011; van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Time is usually considered to be the fourth dimension of the world in which we live, an often essential addition to the three common physical dimensions used to define a point in space, and leading to a more complete notion of 'Space-Time' (DiSalle, 2006; Hawking & Penrose, 2010). Objects, events, and locations in the three spatial dimensions are generally either directly perceived with the human senses of touch, sight, and sound, or are located in a mental map using an internal frame of reference defined by previous experience of those senses (Pasqualotto, Spiller, Jansari, & Proulx, 2013; Lane, Ball, & Ellison, 2015; Kelly & McNamara, 2010). We can identify from which direction a noise has emanated, or which direction an aeroplane in the sky is travelling and, even if the actual magnitudes of extreme microscopic or

interstellar distances are hard to comprehend, they are simply scalar modifications of more manageable distances that have previously been sensed through haptic mechanisms of touch and proprioception, and so can still be readily related to the associated frame of reference (Pasqualotto et al., 2013; Lane et al., 2015; Kelly & McNamara, 2010). Unlike the physical dimensions, however, time itself is completely intangible. Although studies have identified neuronal mechanisms for aspects of time such as interval timing and detection of simultaneity (Merchant & Yarrow, 2016; Murai, Whitaker, & Yotsumoto, 2016; Rohde & Ernst, 2014; Iversen & Balasubramaniam, 2016; Wittmann, 2013), there appears to be no specific sensory organ for time (Bender & Beller, 2014; Gibson, 1975; Srinivasan & Carey, 2010; Matthews & Meck, 2016, 2014; Bausenhardt, 2014), and no evidence that sense organs operate with any notion of absolute time (Vroomen & Keetels, 2010; Keetels & Vroomen, 2012; Bausenhardt, 2014). Poppel (2004) contrasts the ease of defining where something is, by using reliable spatial sensory systems, with the much harder task of specifying when something happened.

The way that time is both of such fundamental importance, and yet intangible, has intrigued philosophers and psychologists alike. Philosophers such as Aristotle linked time with motion in the physical dimensions, and considered that it can only be sensed or measured by means of observable events (Rau, 1953), while in 4<sup>th</sup> century AD, St Augustine said "I measure time in my mind" (Translation quoted from Poppel 2004, p. 296), thus reflecting the practical consideration that, to recognise a pair of non-simultaneous events must require, as well as a sensory mechanism, some form of memory for, at least, the earlier event. Early psychologists, such as James (1890), and Nichols (1891), also considered the relationship between time and events in physical space, "Awareness of change is thus the condition on which our perception of time's flow depends" (James, 1890, p. 261).

A linkage between time and space can be clearly seen in the study of language and metaphor (Nunez & Cooperrider, 2013; Boroditsky, 2011, 2000; Lai & Boroditsky, 2013; Hartmann & Mast, 2012; Casasanto & Boroditsky, 2008, 2003; Sinha & Gardenfors, 2014;

Majid, Gaby, & Boroditsky, 2013; Miles, Betka, Pendry, & Macrae, 2010; Winter, Marghetis, & Matlock, 2015; Torralbo, Santiago, & Lupianez, 2006; Evans, Bergen, & Zinken, 2007) providing many examples of times being described, either in terms of location, or movement, in space. The English language contains temporal location terms such as ‘long ago’, ‘distant future’, ‘Monday to Friday’, and terms involving movement in space, such as ‘summer is approaching’, ‘past its sell by date’ or, ‘the meeting has been brought forward’. For metaphors relating to movement there are also two distinct types depending upon whether the observer is static with the events/time flowing past them (moving time), or the observer is moving through a static arrangement of events/time (moving ego). Early theoretical work by Clark (1973), giving examples of such metaphors in English, has been extended experimentally by many others (Boroditsky & Ramscar, 2002; Matlock, Ramscar, & Boroditsky, 2005; Bender, Rothe-Wulf, Huther, & Beller, 2012; Bender & Beller, 2014), exploring use in other languages (Yang & Sun, 2016; Lai & Boroditsky, 2013), the effects of priming with phrases (McGlone & Harding, 1998; Boroditsky, 2000), with images (Boroditsky & Ramscar, 2002), with physical artifacts (Duffy, 2014), using auditory stimuli and responses (Walker, Bergen, & Nunez, 2014), and in real situations outside of laboratories (Matlock et al., 2005; Boroditsky & Ramscar, 2002). Linkages between metaphors and a wide variety of physical items and concepts (including space, time, and number), were explored from a linguistic point of view by Lakoff and Johnson (2003), leading to a framework of Conceptual Metaphor Theory (CMT) in which it is argued that linkages between metaphors and domains which are in common use, such as arguments, vision, money or time, reflect actual links at a cognitive level that assist in reasoning about that domain. Taking a more neurological approach, Walsh (2003) proposed "A Theory of Magnitude" (ATOM), suggesting that there is a common mechanism within the brain for comparing magnitudes within a range of domains (covering space, time, and number, but also including luminance, sound volume, etc). Some consider these two theories (CMT and ATOM) to be contradictory, whether directly (Bottini

& Casasanto, 2010), from a developmental point of view (Bottini & Casasanto, 2013), or from evidence in animals as well as humans (Merritt, Casasanto, & Brannon, 2010). However, in an overview of the subject, Winter, Marghetis, and Matlock (2015) conclude that the two theories are mostly complementary and suggest that both should be taken into account when examining interactions between space and time.

A number of researchers have studied wider interactions between time, space, and number to provide evidence of underlying common representations or mechanisms. Galton (1880) noted a number of such interactions, including both numbers and dates in visualised positions and sequences. Interactions have been studied between space and number in the SNARC (Spatial Numerical Association of Response Codes) effect identified by Dehaene, Bossini, and Giraux (1993) where, in its most fundamental form, experiment participants responding, using yes/no response sensors, such as push buttons placed to their left and right, generally respond faster when smaller numbers are associated with a ‘left’ response and larger numbers to a ‘right’ response (Deng, Chen, Zhu, & Li, 2016; Viarouge, Hubbard, & McCandliss, 2014; Hubbard, Piazza, Pinel, & Dehaene, 2005; McCrink & Opfer, 2014; Nunez, 2011; Nuerk, Wood, & Willmes, 2005). The extension to linkages between space, and time, in the theories of CMT and ATOM can also be found experimentally, in the STEARC (Spatial Temporal Association of Response Codes) effect, identified by Ishihara, Keller, Rossetti, and Prinz (2008), and also studied by many others (Fabbri, Cellini, Martoni, Tonetti, & Natale, 2013; Fabbri, Cancellieri, & Natale, 2012; Mioni, Stablum, & Grondin, 2014; Vallesi, McIntosh, & Stuss, 2011; Marin et al., 2016; Vallesi, Binns, & Shallice, 2008). Both the SNARC and STEARC are considered to arise from interactions between the activities involved in the experiment and internal representations, for the SNARC effect that internal representation is in the form of an innate mental number line (MNL) (Rugani et al., 2017; Fischer & Shaki, 2014; Winter, Matlock, Shaki, & Fischer, 2015), and for the STEARC effect, a Mental Time Line (MTL) (Bonato, Zorzi, & Umiltà, 2012; Boroditsky, Fuhrman, & McCormick, 2011; Hartmann & Mast, 2012; Arzy, Adi-Japha, & Blanke, 2009;

Droit-Volet & Coull, 2015). Each of the lines represent their respective quantities increasing in a given direction. Some of the MTL research has concerned spatial interactions with estimation of intervals and short duration events (Mioni et al., 2014; Vallesi et al., 2008; Alards-Tomalin, Leboe-McGowan, Shaw, & Leboe-McGowan, 2014; Bizo, Chu, Sanabria, & Killeen, 2006; Takahashi, 2006; Morrone, Cicchini, & Burr, 2010; Roitman, Brannon, Andrews, & Platt, 2007; Ferguson & Martin, 1983; Bonato et al., 2012) while other work has examined longer durations (Weger & Pratt, 2008; Pitt & Casasanto, 2016; Casasanto & Bottini, 2010; Aguirre & Santiago, 2017) more akin to those used in daily life and hence more closely related to commonly used timelines. Frequently researchers have found cultural differences in the orientation and direction of both mental number lines (Winter, Matlock, et al., 2015) and mental timelines (Fischer & Shaki, 2014; Wood, Willmes, Nuerk, & Fischer, 2008; Shaki, Fischer, & Gobel, 2012; Shaki, Fischer, & Petrusic, 2009). In each specific instance, however, the lines involved in the MNL and MTL provide single dimensional baselines against which numbers or times can be compared and hence provide a foundation for the wider uses of the term timeline which often involve two dimensions.

## **Timelines**

The Oxford English Dictionary defines a timeline as: “A graphical representation of a period of time, on which important events are marked” (Oxford English Dictionary, 2012), and recording time in this way has a very long history. Early Greek and Roman scholars would list key names and events in chronological order in books, and in stone carvings etc. (Grafton & Rosenberg, 2010). Individual lists or chronologies can show absolute dates, duration of events, and the relationships between events, in terms of their position within the overall sequence. They can also provide a clear focus on the key items which might otherwise be hard to pick out from a purely textual narrative. When Eusebius of Caesarea, a Christian theologian of the fourth century, combined more than one list in a tabular form, it was then also possible to see inter-relationships, in a temporal sense, between the items in

the different lists. Grafton and Rosenberg (2010, p. 26) refer to his chronicles as providing the “model for later timelines for centuries to come”, while Croke (1982, p. 195) describes it as “one of the most influential books of all time”.

Timelines are used in a wide variety of fields some examples of which can be seen in

Table 2.1

Area of use	Examples
Social network trends	Khurana et al. (2011)
Large scale data	Andre, Wilson, Russell, Smith, and Owens (2007)
Software development	Treude and Storey(2009)
Overview of research topics	Chen, S, Huang, and Chen, D (2012); Zhang, Liu, and Zhao (2008)
Digital Forensics	Inglot, Liu, and Antonopoulos (2012); Olsson and Boldt (2009)
Medical data	Plaisant, Milash, Rose, Widoff, and Shneiderman (1996)
Personal data and family histories	Gemmell, Bell, Lueder, Drucker and Wong (2002); Bennett, Fraser, and Balaam (2012)
Sensemaking in intelligence analysis	Nguyen, Xu, Walker, and Wong (2014)
Producing timelines from text	Fulda, Brehmel, and Munzner (2016)

**Table 2.1** – Examples of Varied Uses of Timelines

## Timelines and schools

One area where the use of timelines is heavily promoted is in the field of education, particularly in the teaching of history. Educationalists of all types, academics, practitioners, and politicians, have promoted the use of timelines for teaching history. Keating and Sheldon (2011, p. 30) illustrate this, quoting from the English Board of Education in 1928 "a few vital dates and facts should, therefore, be driven home at every opportunity - preferably by the use of a time chart". This view has been maintained, with the national Adviser for history in Ofsted (The UK Government Office for Standards in Education, Children’s Services and Skills) saying that "much can be achieved now in improving pupils’

sense of time by teachers regularly using timelines" (Maddison, 2011, p. 9). Meanwhile Thornton and Vukelich (1988, p. 79) point out that the value of using timelines is something that has been asserted, but that these assertions were not empirically based, saying that "Little can be confidently claimed about the effects of time lines on children's understandings", and Burny, Valcke, and Desoete (2009, p. 484) consider such instructional practices in time-related competences to be "driven by ideology, faddism, politics and marketing" rather than by empirical evidence.

A few studies have examined aspects of timeline use in education. Masterman and Rogers (2002) examined the use of interactive multi-media, some of which involved examples of timelines, to scaffold<sup>1</sup> the understanding of chronology by primary age children in an English school. Although they reported positive outcomes, the work was based upon teacher feedback rather than any direct, quantitative, measures. Prangma (2007) studied the effects of graphics (both pictures and diagrams) in support of learning from historical text and associated student collaboration, finding that, in her studies, it was only when the graphics were integrated into timelines that learning (both individual and collaborative), as assessed by an immediate retention test, increased. However, this was seen to be a short term effect and was no longer present at a subsequent retention test around 6 weeks later, and a conclusion is drawn that more work is needed on the understanding and retention of temporal relationships through the use of timelines. Korralo (2010b) examined the use of virtual environments to improve learning of historical chronology, and found that learning could be improved at the immediate testing stage, although, again, delayed retention was not significantly different from a standard text based teaching approach. There was also evidence that the engagement of the spatial memory processes involved in navigating virtual reality environments could increase the amount learned. Extending the work to include multiple timelines, Korralo, Foreman, Boyd-Davis, Moar, and Coulson (2012) found

---

<sup>1</sup>Scaffolding is a widely accepted theory in developmental literature that describes how new topics are learned by building upon existing concepts. Developed by Wood and Bruner in the 1970's (Bruner, 2006), it is also often linked to Vygotsky's Zone of Proximal Development (Chaiklin, 2003).



benefits for university age students in being able to better remember relationships between events and to gain a "Big Picture" from virtual reality-based examples of timelines. Most of these studies have concluded that more research is needed regarding the use of timelines in learning and, taken together with the high expectations of educationalists regarding their benefits, as described earlier, it seems clear that there are significant gaps in knowledge in this area. The most fundamental of these gaps is that politicians and educational professionals are working to encourage the use of timelines in the curriculum with little clear evidence to support the benefits. It is suggested that this comes from a long history of essentially anecdotal experiences (Keeting, 1976; Maddison, 2011), and therefore the first major gap to address is whether timelines can be empirically shown to improve learning. Where research has started to explore this gap, it has been limited by its complexity. Prangma (2007) and Korallo (2010b) both found graphics and virtual environments could help in the initial understanding of material, but neither showed an impact in long term retention. Furthermore Prangma considered collaboration alongside graphics, while Korallo used novel virtual environments. Therefore the second objective is to build on their work, simplifying the question and investigating whether timelines are better than text information without including the additional aspects of collaboration or novel virtual environments.

### **The Process of Learning**

Learning is considered by Olson (2015) to be a very important topic in psychology, but one that is also very hard to define. One starting point is the definition given by Kimble (1961, p.6) "Learning is a relatively permanent change in a behavioral potentiality that occurs as a result of reinforced practice". This has proven relatively popular (Houston, 2014; Kimble, 1964). However Olson (2015), while acknowledging its popularity, notes that it is not universally accepted by psychologists, and suggests that it should simply be considered a starting frame of reference for more detailed definitions. He does, however, also note a

wider common public definition, that learning consists of "'Comprehension, 'knowledge,' or 'understanding' gained through practice and experience" (Olson, 2015, p.1). This, more general, description aligns well with a cognition focused definition used by Paas and Sweller (2014, p. 41), "Any change in long-term memory involving an accumulation of information", which is the basis of the definition adopted in this thesis. The requirement for 'accumulation' matches the term 'gained' in the earlier, common, view and ensures that it is an increase in knowledge/skill that is being examined and not any loss through ageing or disease for example.

There are also different stages or depths of learning to be considered. The first stage involving an ability to recognise an item, image, or pattern, when it arises subsequently (Malmberg, 2008; Freeman, Heathcote, Chalmers, & Hockley, 2010; Besson, Ceccaldi, Didic, & Barbeau, 2012). At a deeper level such items, images, or patterns can be actively recalled at a later stage in response to an indirect prompt or cued recall (Aue, Criss, & Fischetti, 2012; Mulligan & Picklesimer, 2012; Salthouse, 1982). Having an understanding of the meaning of the item, image, or pattern is a yet deeper step while the deepest level involves the notion of transfer (Barnett & Ceci, 2002) where the knowledge and understanding can be correctly applied to a different situation. For example learning about percentages in mathematics at school, and then being able to apply this while shopping when there is a sale on.

Within this thesis, the term comprehension is used in the sense that it often is in schools, where the students are assessed on their ability to extract relevant information, and answer specific questions on material that they have been given. In schools, students are often asked to summarise sources, or to answer questions based on such reading material. Thus comprehension is considered to be the ability to use materials to find the answer without necessarily understanding the complete content in depth.

## **Critical characteristics of the cognitive system underlying learning**

Many models of human cognition, memory, and processing, have been proposed. The various models aim to match and explain empirical observations/measurements, clinical neurological examinations (both by surgery and by brain scanning), and pharmacological effects. Some of these operate at the neuronal level and attempt to build “bottom up” models (Eliasmith et al., 2012). Most however, aim to fit “top down” models with the various observed effects. In early work on a frequently cited theory of memory, Bartlett (1932) proposed that learning involves the creation of "Schemas", patterns that can be subsequently recognised or performed as patterns of activities. In another classic study, Miller (1956), reinforced in later work by Simon (1974), explored some limits to memory capacity, and proposed that effective capacity is increased when discrete items are associated together in "chunks". Miller's paper focused upon short term or "immediate" memory, while Bartlett (1932) was more concerned with longer term memories. Their concepts of "schemas" and "chunks" have some strong similarities however, since both relate to grouping items together. Most of the memory and processing models in popular use take account of observable limits to the number of items that can be handled simultaneously while being incorporated into schemas or chunks, and use a notion of short term or ‘working’ memory where connections are made, both between items in that memory, and also with items in long term storage that can then assist in their later retrieval. The models differ in whether they consider this short term/working memory to be distinct from the long term storage (Atkinson & Shiffrin, 1968; Baddeley, Kopelman, & Wilson, 2004; Baddeley & Hitch, 1975; Baddeley & Hitch, 2000) or whether the underlying memory resources are the same, but the level of activation/depth of recoding is the critical differentiator in respect of short/long term retention (Cowan, 1988, 2014, 2001; Craik & Lockhart, 1972). The initial notion of levels of processing (LOP) proposed by Craik and Lockhart (1972), that the deeper the initial processing taking place during the learning, the greater the retention that will result, is considered an important framework (Craik, 2002; Loaiza, McCabe, Youngblood,

Rose, & Myerson, 2011; Ekuni, Vaz, & Bueno, 2011), with, for example, applications in learning from combinations of text and maps (Scevak & Moore, 1998). Extending the framework to take account of the expected style of testing for recall, resulted in the concept of Transfer Appropriate Processing (TAP) suggested by Morris, Bransford, and Franks (1977), where the format of the learning activities, and the associated material information should be matched to the expected recall tests in order to result in the strongest retention (as measured by the particular test). Many of the studies that have shown the effects of both LOP and TAP, have been word based (Mulligan & Picklesimer, 2012; Morris et al., 1977; Craik & Lockhart, 1972), but those examining images (Intraub & Nicklos, 1985), diagrams (Lowe & Boucheix, 2008), and map/text combinations (Scevak & Moore, 1998), could have relevance in the context of timelines.

There are critics of the detail of each of the memory and associated processing models; Nairne (2002) points out experimental contradictions with simple theories of short term memory, while Jones, Macken, and Nicholls (2004) argue that short term memory need not be a separate construct, but can instead be explained as an artefact of process. However, overall, there appears to be a consensus regarding both the medical and empirical evidence, that it is possible to differentiate between not only the retention time of short term and long term memories, but also their relative storage capacities (Unsworth & Engle, 2007; Ma, Husain, & Bays, 2014). It is these limitations that impact upon the bandwidth or rate at which learning that can take place and form crucial foundations for a number of learning theories that seek to optimise learning.

### **Dual Coding Theory (DCT)**

In studies relating to learning and recall, and to verbal and visual representations Paivio (1991) developed a Dual Coding Theory (DCT) which postulates that verbal and non-verbal (usually visual) information is both handled, and stored, separately, and that each can provide a trigger for the recall of the other (Clark & Paivio, 1991). These internal mental

images, and verbal symbols, which he termed "logogens" and "imagens" respectively, can be produced by the various sensory inputs: auditory, visual, haptic, and motor properties - with different imagens and logogens for each mode. Although independent, the two systems interact so that each can activate the other and can therefore have additive effects, resulting in increased power of recall. A picture with a clearly identifiable subject that can be named, or a word that also brings to mind a clear picture, are both more memorable than an abstract picture, or a concept word with no obvious related image. The way that concrete nouns can bring to mind associated images also provides a possible explanation for the Picture Superiority Effect (Shepard, 1967; Paivio & Csapo, 1973) where images are generally recalled better/faster than words because the pictures are coded twice (or more strongly in a depth of processing model), once as the visual version, and once as the verbal version. It should be stressed that the dual coding, as described above, relates to the processing channels and not necessarily to the sensory channels involved. A spoken word may conjure up an image, while also being processed as a verbal symbol. A word, read in text form, and so using only a visual sensory channel, could also produce both its verbal symbol and an associated image.

Paivio and others have continued research to support and widen the scope of the theory (Paivio, 2014; Paivio & Sadoski, 2011), examining bilingual aspects (Paivio, Clark, & Lambert, 1988; Paivio & Lambert, 1981), physiological evidence (Welcome, Paivio, McRae, & Joannis, 2011; Kounios & Holcomb, 1994), and extension to other sensory modes, of sound (Crutcher & Beer, 2011; Bolls, 2002), and smell (Lwin, Morrin, & Krishna, 2010). The further research has also provided context or contrast for related work (Sadoski, Paivio, & Goetz, 1991; Paivio & Sadoski, 2011; Paivio, 2013; D'Agostino, O'Neill, & Paivio, 1977) and, of particular relevance to the studies here, its application to education (Paivio, 2008), text (Sadoski, Kealy, Goetz, & Paivio, 1997), and maps (Kulhavy, Stock, Woodard, & Haygood, 1993; Diana & Webb, 1997) where the work incorporated the DCT into a specific hypothesis of Conjoint Retention (Kulhavy, Lee, & Caterino, 1985)

which is discussed later.

The DCT has frequently been used to justify the use of pictures combined with text in learning materials. However, when considering a wide range of diagrammatic representations, Schnotz and Bannert (2003) point out that Paivio's dual coding theory can be too simplistic, and that both the form of graphical representation, and the prior knowledge and ability level of the reader, are very important considerations. In some cases the addition of graphical representations can actually interfere with the learner's formation of a mental model. It should also be noted that, whilst the DCT and associated research (D'Agostino et al., 1977; Sadoski & Paivio, 2004; Carney & Levin, 2002; Paivio & Csapo, 1973), do discuss differences between concrete and abstract words, the majority of the studies involved still use pictures of identifiable objects rather than abstract pictures (where there may be much less of a connection with an equivalent textual representation). Timelines may incorporate images (see for example Figure 2.1) or be more plain in style (see Figure 2.2), but should be considered, in any case, as images or graphics in their own right. The Oxford English Dictionary (2012) certainly considers them as graphics in the following definition of a timeline: "A graphical representation of a period of time, on which important events are marked". Aigner, Miksch, Schumann, and Tominski (2011) take a similar view of them as graphical representations when describing their history and use. Prangma, Boxtel, Kanselaar, and Kirschner (2009) noted that images can be concrete or abstract. Timelines are an excellent example of an abstract image, where the layout of the text information in a chronological and graphical style results in an image or "big picture" which adds value to the text. In a timeline, not only does the date convey when something happened, its position within the timeline also gives a visual reminder. Furthermore timelines provide additional context and a visual ordering of events, which can be much harder to convey, and read, when in a text form. Some timelines, however, use pictures, cartoons, or pictorial symbols to mark events or provide linked information, and so become more of composite image mixing concrete and abstract. However, whether a single abstract image, or including concrete

sub-images, it is acceptable to consider the overall two dimensional timeline as an image, especially when considered in contrast to written text. The DCT therefore should have relevance in studies concerning the cognitive effects of timeline representations. Moreover, if timelines are considered to be "maps of time" (Salomon, 2012; Christian & McNeill, 2011), then the more specific application of the DCT to learning from maps in the Conjoint Retention Hypothesis (Kulhavy et al., 1985) could also have some relevance.

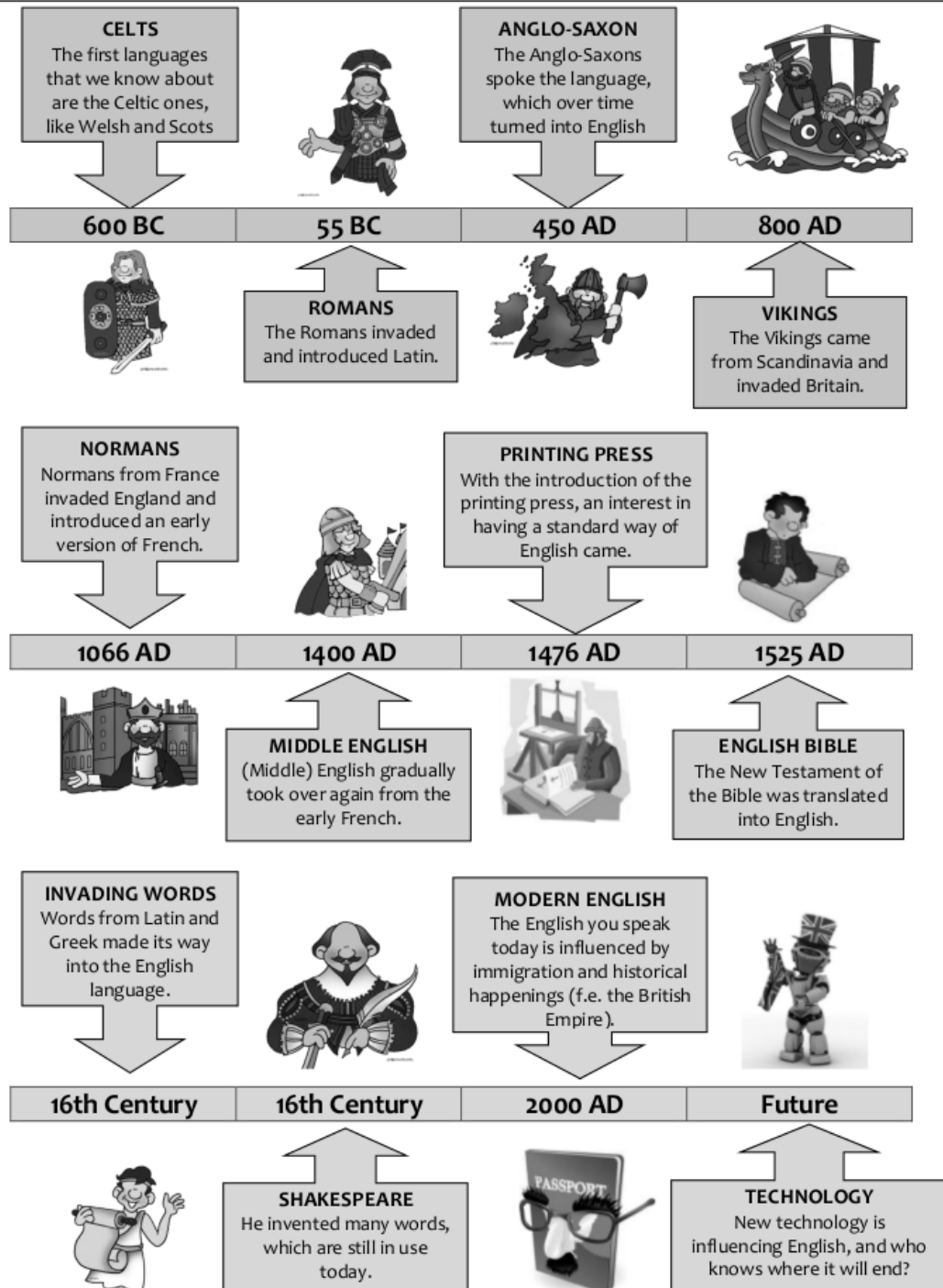
### **Conjoint Retention Hypothesis (CRH)**

The Conjoint Retention Hypothesis (Kulhavy et al., 1985) is essentially a specific example of the DCT applied to learning where both a graphical map and directly associated text are provided. Early work by Dean and Kulhavy (1981) examining increases in comprehension and retention of information when a map was used in conjunction with text, led to further examination (Abel & Kulhavy, 1989), some of which suggested that to be effective, the maps needed to have a distinct structure (Abel & Kulhavy, 1989) and be shown before the text (Kulhavy, Stock, & Kealy, 1993; Verdi, Johnson, Stock, Kulhavy, & Whitman-Ahern, 1997). Examination of the hypothesis by other researchers failed to replicate the results however, finding no difference against learning from lists (Griffin & Robinson, 2005; Ungar, Blades, & Spencer, 2000). Most of the studies concerned geographical maps, but some (Newbern, Dansereau, & Patterson, 1997) extended the work to spatial-semantic displays such as concept maps of text, and to knowledge maps/mindmaps, both of which are structures more akin to timelines, although pointing out the need for further research in the area, as does Vekiri (2002). Although there is an attraction to viewing a timeline as a "map of time", benefiting from any conjoint retention, the first priority in this thesis is to address the role of timelines as graphic elements of multimedia in the CTML.

Both dual coding theory and the conjoint retention hypothesis link increased learning to multiple channels or routes to the memory. However, effective learning also needs to ensure that those channels are not overloaded with information and the appropriate design of

# The History of the English Language

Where did the English we speak today come from?



Source: [http://www.childrensuniversity\\_manchester.ac.uk/media/services/thechildrensuniversityofmanchester/flash/timeline.swf](http://www.childrensuniversity_manchester.ac.uk/media/services/thechildrensuniversityofmanchester/flash/timeline.swf)



learning materials needs to take into account the associated cognitive loads.

**Cognitive Load Theory (CLT)** While studying the problem solving process, and comparing approaches taken by experts with those of novices, Sweller (1988) developed Cognitive Load Theory (Sweller, 1988, 1994; Sweller & Chandler, 1994; van Merriënboer & Sweller, 2005; Paas, Renkl, & Sweller, 2004; Sweller, 2010). In this it is argued that materials used for learning (text, diagrams, slides, etc.) result in three identifiable loads on the learner's cognitive system. An "intrinsic" or "essential" load that reflects the fundamental information content of the item/concept being learned, an "extraneous" load reflecting the effort that is needed to understand the particular way in which the learning item is presented (for example, distracting borders around a diagram can unduly increase this load) and the "germane" or "generative" load, the cognitive load associated with making the necessary connections to understand and store the item. His work on problem solving examined the different ways in which an expert can "short circuit" some of the steps that novices have to go through, in order to reach a conclusion much more rapidly. He built upon the concept in Miller's work that there is a limit to working memory capacity, highlighted the heavy cognitive load that novices undergo when solving problems, and suggested that, in some cases, this must overload their cognitive processing capacity so that there is no capacity left to transfer the learning into long term memory (Sweller, 1988).

Considering the DCT and the CLT, it is possible then that timelines could be beneficial for learning as they would trigger dual coding through the visual and verbal pathways, and may maximise the germane load through helping to visualise relationships. However there are many ways that inappropriate, irrelevant, or duplicate information on the timeline may also act as a distraction that could produce additional extraneous load on the learner, overloading their capacity, and thus preventing successful learning. Appropriate construction of timelines therefore requires an understanding of how to minimise these extraneous loads through an overall understanding of how learning from multimedia works.

## **The Cognitive Theory of Multimedia Learning (CTML)**

While it is possible to conceive of a timeline without any text at all, perhaps one showing the order of events by just using images of those events on a line, in all common cases a timeline will also include some text, at least in the form of axes, possibly a key, and labels/descriptions of events. In a practice guide produced for the United States Department of Education, Pashler et al. (2007) consider that the incorporation of diagrams, with associated prose, into multimedia learning materials, can provide tangible learning benefits. Finding there to be a 'moderate' level of supporting evidence (studies where either external validity or internal validity, but not both, are high, with the corresponding validity moderate), they cite an example, from mathematics, of the incorporation of a graphical number line into instructional material (Moreno & Mayer, 1999b). The Cognitive Theory of Multimedia Learning (CTML) developed by Mayer (2005), and supported in his many subsequent studies (Mayer, 2009a; Massa & Mayer, 2006; DeLeeuw & Mayer, 2008) would consider the incorporation of a timeline into a history text as an example of multimedia. The definition used (Mayer, 2014a, P.2) being "Presenting words (such as printed text or spoken text) and pictures (such as illustrations, photos, animation, or video)". The CTML has three major foundations: Sweller's Cognitive Load Theory (Sweller, 1988, 2010), Paivio's Dual Coding Theory (Paivio, 1991; Clark & Paivio, 1991), and previous work by Mayer (1996) and Wittrock (1992), concerning active learning and organising/integrating new information with existing knowledge. In essence the theory argues that the dual channels implied by DCT, with their processing limitations defined by CLT, will be most efficiently used for learning when both channels are used at, or near, capacity and when the material to be learned involves suitable multimedia representations to bring this about, and is in a form that encourages linkages to be made, both between the representations, and with existing knowledge. The CTML takes these ideas as foundations, and derives a wide range of different principles for instructional design from them, each tested via research. It should be noted that these principles have been adapted and amalgamated multiple times in the

literature, and are not seen to either be mutually exclusive, or an exhaustive set of principles. Instead they are proposed as guidelines which may help when using multimedia in learning. The principles within the CTML have been experimentally explored by Mayer (2014a) and many others (Frey & Sutton, 2010; Adesope & Nesbit, 2012; van Genuchten, Scheiter, & Schuler, 2012; Eilam & Poyas, 2008; Lee, Hsiao, & Ho, 2014), in a variety of ways, and a comprehensive summary of the work is provided by *A Handbook of Multimedia Learning* (Mayer, 2014a). During the time between the first steps towards a comprehensive theory (Mayer & Gallini, 1990; Mayer & Sims, 1994) and the publication of the Handbook, the instructional design value of some of the principles defined within has been increased by the inclusion of greater definition of the boundaries of applicability and identification of the way that, in some cases, outside of those boundaries, the effects may reverse and become counterproductive.

When considering timelines as an example of multimedia it is clear that some of these principles are more relevant than others. This discussion takes the opportunity here to reflect on some of the relevant principles and how they should be applied to timelines. More detail, including references to related research, can be found in Tables A.1 and A.2 on page 236.

**Multimedia principle** Learning is increased when the materials used contain words and pictures rather than just words. Timelines are seen to be a type of picture or map. They allow the user to see the "big picture" and can (based on DCT and CRH) help people to remember information as they may be able to encode the information twice, first as a verbal/text through reading the content, and secondly (perhaps more importantly for gaining understanding of the historical context) the ability to recall the whole picture. Remembering that a name appeared on the right hand of the page, may help a student to put events in order even if they cannot remember the exact date that it occurred.

**Split Attention** Learning is increased when learners do not have to split their attention between several sources of information - Well designed timelines follow this principle, as

they can act as a summary document for a wide range of resources. By viewing each small event as a large "picture" the student does not have to read through many small paragraphs about events, but instead sees how they link or interact through the chronological map of a timeline. This principle has two closely related principles, that of **Spatial Contiguity**- the learning is increased when text and associated graphics are physically close together and that of **Temporal Contiguity** - People learn better when corresponding words and pictures are presented simultaneously rather than in succession.

**Redundancy** Learning is reduced when the same information is provided in multiple forms or is over elaborate in extraneous detail. As timelines are often attempting to display a large amount of chronological information in a relatively small area, they need to be both concise and precise. As a picture can be worth 1000 words, a timeline could also be said to be the equivalent of 1000 words.

**Signalling/Cueing** Learning is increased when key information is pointed out by visual cues such as colour coding or bold text. Timelines often feature clear headings, borders, or the use of colour to distinguish between events. Sometimes themes can be highlighted in the timeline to add a further dimension (e.g. for a timeline showing reasons leading to the second world war, economic, social, political and religious events may be colour coded differently).

**Coherence** Learning from multimedia lessons can be increased when the graphics are relevant to the goal (and conversely decreased when they are not). Care must be taken that unnecessary detail is not included if that may distract the learner as a "seductive detail" or confuse the learner and hence result in what Sweller (2010) termed, extraneous load.

**Generative Drawing** Encouraging learners to produce their own diagrams while learning can increase learning (Van Meter & Garner, 2005; Van Meter, Aleksic, Schwartz, & Garner,

2006; Lin et al., 2017; Schmeck, Mayer, Opfermann, Pfeiffer, & Leutner, 2014). This is included here as teachers often encourage students to build their own timeline based on multiple sources of information.

### **The absence of timelines as examples of CTML**

There are many more principles in the CTML but these are considered the most important for timelines. It should be noted, however, that Mayer (2014b) does not address timelines directly, nor does any of the related research; indeed Ayres (2015) points out that the vast majority of studies have involved STEM (Science, Technology, Engineering, Mathematics) subjects, and that most of the research has only been carried out in laboratory conditions. Outside of these STEM based studies, a language based study is discussed by Kalyuga and Sweller (2014), an accountancy based study by Blayney, Kalyuga, and Sweller (2016, 2010) and a history based study (which was focused upon historical enquiry and not timelines) by McNeill, Doolittle, and Hicks (2009). Furthermore the diagrams used in many of the CTML studies are also generally in the form of illustrations (the workings of a car braking system, or an air pump) or directly representational (the physics behind a lightning strike, or a schematic diagram of a pulley system) and so these can be considered as concrete images. In contrast, a timeline usually provides a more abstract image. There is therefore a significant gap in the CTML literature regarding whether the principles can be effectively applied outside of the STEM subjects, and to abstract rather than concrete images.

### **An Integrated Model of Text and Picture Comprehension (ITPC)**

The CTML is not the only model of multimedia learning, Schnotz and Bannert (2003) also examined multimedia learning, with DCT and CLT as foundations (Schnotz, Ludewig, et al., 2014; Schnotz & Kurschner, 2007; Schnotz, Mengelkamp, Baadte, & Hauck, 2014) resulting in an Integrated Model of Text and Picture Comprehension (ITPC). In the ITPC, however, the notion of visual and textual/auditory channels is further extended to take

account of research into text comprehension (Kintsch & Van Dijk, 1978; Graesser, Millis, & Zwaan, 1997) which suggests that readers/listeners produce three distinct mental representations; an initial surface representation that simply reflects the words used (allowing them to be repeated even if not understood), a propositional model that reflects the individual ideas contained in the words, and finally, at the deepest level, a mental or semantic model of the content. Schnotz and Bannert (2003) propose that graphical information is also assimilated at those levels, so, for example, a diagram such as a timeline may be comprehended at the surface level merely as lines on a page, in more depth as showing the relationship between events, and, at the deepest semantic level, as the history that it is conveying (Schnotz & Baadte, 2015). Unlike the CTML, where the text and image representations are considered as separate until eventually integrated into long term memory, the ITPC suggests that both the propositional and the semantic models are based on all modes of information (also incorporating relevant items from long term memory). While there are differences between the CTML and the ITPC, they are not relevant in the context of this thesis, as both suggest that multimedia will be beneficial to learning, albeit by different methods, and this thesis is interested in whether timelines can improve learning, not exactly how the learning has occurred. As the CTML has had more research based upon it, and the principles are more clearly defined, it is appropriate in this work to consider the CTML as the guiding method for this body of work, and to use it to structure the experiments and the understanding of how timelines can support the learning of historical concepts.

### **Estimating Cognitive Load**

When assessing the effectiveness of different types of educational material in order to develop guidelines for creation of the most effective content, it is necessary to measure, or at least estimate, the cognitive load involved since material that overloads the cognitive capacity of the learner will not be fully retained. There are a number of ways in which cognitive load can be assessed. Generally cognitive load assessment methods can be

divided (Brunken, Seufert, & Paas, 2010) into: subjective ratings (Hart & Staveland, 1988; Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Leppink, Paas, Van der Vleuten, Van Gog, & Van Merrienboer, 2013; Hart & Staveland, 2005), performance measures such as accuracy and response times (Van Zandt & Townsend, 2014), including performance of dual tasks (Brunken, Plass, & Leutner, 2003; Piolat, Olive, & Kellogg, 2005), behavioural measures such as voice changes (Khawaja, Chen, & Marcus, 2014), and physiological measures such as: Galvanic Skin Response (Nourbakhsh, Wang, Chen, & Calvo, 2012), heart rate (Mulder, 1992; Cranford, Tiettmeyer, Chuprinko, Jordan, & Grove, 2014), and pupil dilation/blink rate (Zekveld, Heslenfeld, Johnsrude, Versfeld, & Kramer, 2014; Chen & Epps, 2013). Subjective measures are commonly used since they are relatively easy to incorporate (van Gog & Paas, 2008; Paas et al., 2003). Early work, adapting scales used in general workload assessment, by Paas and Van Merrienboer (1993) resulted in the subjective cognitive load measure (SCL) using a seven or nine point Likert scale, and which Kalyuga, Chandler, and Sweller (1999) considered to be both sensitive and reliable. This has been refined (van Gog & Paas, 2008) and used by many others (Ayres, 2006; Kuhl, Scheiter, Gerjets, & Edelman, 2011; Lin et al., 2017), although Brunken et al. (2003) highlight its broad nature as being unable to differentiate between task difficulty, individual competency or different levels of attention. Another popular scale, originally developed for assessing pilot workload, the NASA - TLX (Hart & Staveland, 1988) has also been adapted for use in assessing cognitive loads in learning (Paas, Ayres, & Pachman, 2008; Mayes, Sims, & Koonce, 2001). Wiebe, Roberts, and Behrend (2010) compared the TLX and the SCL, finding similar results at high levels of intrinsic load (the load due to the subject of study itself, rather than any aspect of the presentation), but higher sensitivity in the SCL at lower levels of intrinsic load. This latter finding, together with the overall popularity of the SCL in the field of multimedia research, Paas et al. (2003) for example, noted 24 studies that had used it, suggests that it is the measurement to use in studies of timelines as examples of multimedia.

It should be noted, however, that a number of the cognitive load measures, including the SCL, rely on interrupting the learner to establish the load they are undertaking. While some work has used biological measures, these too can interrupt the naturalistic experience and act as an additional confound when considering the translation of studies in the laboratory to real world environments. It has been suggested that response times could be used as an accurate method for measuring the level of load a learner is experiencing (Chen et al., 2016), but there is limited research in this area. Therefore a key gap in the cognitive load research is understanding whether response times can be used as an alternative to the more commonly used SCL.

## **Learner preferences, spatial abilities, and other factors for consideration**

In addition to multiple theories about learning and associated loads, it is also important to consider other factors which may influence the ability to study, store and reflect upon information so that it can be learned. This next section provides an overview of the key aspects that should be considered when designing studies to address the earlier highlighted research gaps.

**Learner Preferences** During learning there are two fundamental elements that have to be considered, the materials used and the learner themselves. The idea that learners have preferred styles of learning such as visual, auditory, verbal, and kinesthetic, and that learning can be improved by ensuring that the learning materials and approach are adapted to the learner's preference, has been popular in education for many years. Indeed whole industries have built up based upon assessing learning styles and suitably adapting materials. Pashler, McDaniel, Rohrer, and Bjork (2008), referring to a "thriving set of commercial activities" (p. 106), point to a study describing 71 different schemes, and conclude that there is no strong, methodologically sound, evidence to support any of the theories. A similarly forthright view was put forward by Kirschner (2017) in a paper entitled "Stop propagating



the learning styles myth". The absence of a statistically sound link between preferences and performance has also been highlighted by a number of others (An & Carr, 2017; Cuevas, 2015; Rogowsky, Calhoun, & Tallal, 2015; Abdul-Rahman & du Boulay, 2014; Nixon, Gregson, & Spedding, 2007). In a large scale synthesis of meta-analyses covering a wide range of possible influences on academic achievement, Hattie (2008, p.195-197) examined eight meta-analyses, covering 411 studies, but questioned the validity of a number of these, and concluded, with reference to an extensive study by Coffield, Moseley, Hall, and Ecclestone (2004) (which contained many criticisms regarding overstatement, limited impact, and a focus on commercial advocacy), that while learning strategies, and promotion of enjoyment of learning, had benefits, the same could not be said for learning styles. A related review of research reported by Sutton Trust (2018) also considered there to be limited evidence for the effectiveness of adapting teaching and learning materials to learning styles and judged any impact to be low. Pashler et al. (2007) and Mayer and Massa (2003) did, however, find links (albeit without statistically strong effects) between preferences and performance in some specific examples of multimedia learning (Massa & Mayer, 2006), and a few other researchers have also found evidence to support the theories (Hoffler, Koc-Januchta, & Leutner, 2017; Haciomeroglu & LaVenja, 2017), especially when the preferences are measured multidimensionally. None of these has examined timelines in any way however, and, since the effects may differ for this type of material, the inclusion of suitably targeted assessments of learner preferences could be justified in studies of learning through timelines as an exploration of their relevance and applicability for the optimisation of learning processes and materials in the situations examined.

**Spatial Ability** The improved memory for the combination of maps and text that results from conjoint retention theory has been linked to engagement of spatial memory and associated abilities (Robinson, Robinson, & Katayama, 1999) and similar links have been found in studies concerning the understanding and retention of diagrams and texts (Miyake,

Friedman, Rettinger, Shah, & Hegarty, 2001; Pazzaglia & Moe, 2013; Thorndyke & Stasz, 1980; Hochpochler et al., 2013; Sanchez & Wiley, 2014). Of the five main elements of spatial ability summarised by Carroll (1993), two were identified as having the strongest relevance in a meta-analysis of visualisation experiments by Hoffler (2010). These were “spatial visualisation” and the connected measure of “spatial relations”. Many studies of multimedia learning have examined these factors (Massa & Mayer, 2006; Bednarz & Lee, 2011; Pazzaglia & Moe, 2013; Brunken et al., 2003) and hence further experiments extending multimedia learning to other materials, such as timelines, should incorporate both spatial relation measures for comparison. Commonly used measures in this regard are mental rotation tests (MRT): (Shepard & Metzler, 1971; Hegarty, 2004; Miyake et al., 2001; Pazzaglia & Moe, 2013), and object location memory measures (OLM): (Peebles & Jones, 2014; Haladjian & Mathy, 2015). In addition Baddeley and Andrade (2000) consider that vividness of imagery interacts with long term memory and hence assessment of abilities in this regard should also be included in investigations of multimedia learning. It should be noted that these spatial relation measures relate to the multimedia approach involved here, rather than the wider spatial memory aspects investigated in studies using virtual reality (Korralo et al., 2012; Korralo, 2010b; Foreman, Boyd-Davis, Moar, Korralo, & Chappell, 2008).

**Age and development** Age has frequently been shown to be a significant factor in visual, spatial, and text processing memory/abilities (Borella, Ghisletta, & de Ribaupierre, 2011; Borella, Meneghetti, Ronconi, & De Beni, 2014; Meneghetti, Fiore, Borella, & De Beni, 2011; Salthouse, Babcock, Skovronek, Mitchell, & Palmon, 1990), with significant age related declines in older adult performance, in all of these. A large scale Internet based study by Brockmole and Logie (2013) suggested that visual working memory abilities peak around age 20, declining almost linearly thereafter. Studies of multimedia learning with adults should therefore examine age as a potential factor in any performance variation. For

any studies of timeline use in schools, as well as age, there is also a related developmental aspect to be taken into account which concerns the level of understanding of history/time itself. Piaget (1969), considered time to be a particularly difficult concept for children to grasp and this view, through its incorporation into an influential report<sup>2</sup> in England "Children and their Primary Schools" (CACE, 1968), had a long-term effect upon history teaching in the UK (Thomas, 1990). The view that history could not be successfully taught until adolescence held sway until the late 1980s and beyond. Hoodless (1996) and Zaccaria (1978) considered that students, on average, only develop the ability to operate at the formal level of thinking between 16.5 and 18.2 years, and that this is a full 4 years later than the same stage (on average) in mathematics and science. More recent studies by historians, educators, and psychologists (Friedman, 1982; Dawson, 2004; Hoodless, 1996; Stow & Haydn, 2000b), that have explored children's understanding of time in a variety of ways, have however, departed from these Piagetian views. In partial explanation Hoodless (2002) points to today's children having much greater exposure to travel, media, IT, etc., resulting in a much better understanding of time. Her research with children aged between 3 and 9 years, and using conversations about stories as the material for the analysis, showed that, in the right context, use of temporal terms and ability to sequence chronologically is often present at these ages. Burny, Valcke, and Desoete (2009), quoting from a variety of studies where children show significantly greater understanding of temporal matters at much earlier ages than Piaget would suggest, conclude that the concepts put forward by Piaget, although still useful when considering development, are incomplete. They also highlight both the importance of instruction in temporally related competences, and the need for a systematic approach to such instruction, while observing that there is very little related research available to show how to do this.

**Motivation** Learning is an active process in which learners need to engage with the learning material, and their motivation to do this can also have a major bearing on

---

<sup>2</sup>usually referred to as 'The Plowden Report' after Lady Plowden who chaired the committee involved.

comprehension and retention (Leutner, 2014; Heidig, Muller, & Reichelt, 2015; Mayer, 2014b).

Prangasma et al. (2009) noted that learning had three components, Effectiveness (accuracy), Efficiency (speed of learning and retrieval), and being Pleasant to the learner (which can be interpreted as the materials being engaging, or motivating). Similarly Palmer (2005) notes that “Motivation is a prerequisite and co-requisite for learning”. Work by Vygotsky and Piaget also placed the emphasis on the learner, requiring them to self-motivate and explore materials, either through the provision of taught experiences (Piaget) or through collaboration and interaction (Vygotsky). It is clear that motivation is required to learn successfully, however, it is not always clear what the motivation is. Researchers often discuss intrinsic or extrinsic motivation. Intrinsic motivation is being motivated purely by interest and enjoyment, for example playing along with a quiz on the television, whereas in contrast, extrinsic motivation has an obvious reward; to continue the example, the quiz show contestant will have a monetary motivation. Research has suggested that intrinsic motivation is better for learning, with extrinsic motivation sometimes being detrimental to the learning process (Deci, Koestner, & Ryan, 2001).

Building upon the foundations of the CTML, Moreno (2006) proposed the cognitive-affective theory of learning with media (CATLM) which takes account of motivational factors, but observed that existing research gave only limited support to the theory and was mostly laboratory based. Subsequent studies (Park, Plass, & Brunken, 2014; Park, Flowerday, & Brunken, 2015) showed that a number of motivational factors could affect learning, and that a balance needed to be found when designing multimedia materials, between the inclusion of decorative illustrations to engage the learners’ interest, and the need to minimise distraction and maintain coherence, as promoted by the CTML (Mayer, 2014a; Harp & Mayer, 1998).

In this thesis it is anticipated that the timelines will increase motivation by making the learning material more accessible and interesting to the learner. However, as motivation in

learning is a large area of research in its own right, it would not be possible to do it justice in this thesis alongside measuring learning. Instead it will be noted that timelines may be expected to improve learning through motivation, but it will not be measured or explored as a key aspect of this work.

## **2.3 Summary and Research Context**

From the literature discussed above a number of conclusions can be drawn. There are many proponents of the use of timelines to teach history (Maddison, 2011; Hodkinson, 2011; Pickford, 2011), but little has been done to investigate their effect either in laboratory experiments or in schools (Thornton & Vukelich, 1988; Burny et al., 2009; Stow & Haydn, 2000b). Timelines are examples of multimedia and, when their effects on learning are studied, should be considered in terms of the relevant aspects of the CTML (Mayer, 2014a). Most of the research relating to the CTML has used scientific diagrams or pictures as the primary material (Ayres, 2015) and no examples have been found where timelines were used. Furthermore, consideration of timelines as maps of time may also engage spatial memory through the CRH (Kulhavy et al., 1985). Foreman et al. (2008) did explore the connection with spatial memory in respect of virtual reality based timelines, and Prangma (2007) studied the use of multi-modal representations in collaborative history learning, but no studies have been found that examine differences between timeline and text versions of history related information, in both learning (comprehension and retention), and in associated cognitive load.

The various principles within the CTML are based upon an aim of making the most effective use of cognitive load during learning and, whilst there are a number of methods of measuring cognitive load, including the measurement of response times (Van Zandt & Townsend, 2014), CTML related research frequently uses a self report method the SCL (Paas et al., 2003). Subjective measures have been compared to response times for

secondary tasks (DeLeeuw & Mayer, 2008), and comparisons have been made between different self report scales (Wiebe et al., 2010), but no comparisons have been found between any instruments using self reported measurements and response times for the primary tasks.

The key gaps and questions that have arisen in this literature review are therefore:

- Are timelines as effective as politicians and education advisors suggest, in particular is information presented in a timeline format more beneficial than simple text when people are learning?
- Furthermore, are timelines effective in terms of both comprehension and retention aspects of learning?
- Can the CTML be extended beyond STEM to consider timelines as examples of abstract graphics?
- Do response times correlate to self reported measures of cognitive load (SCL) ?
- Are the potential benefits of timelines limited to participants with certain characteristics e.g. spatial ability, age or learner preference?

The gaps described above led to the definition of a set of hypotheses, and studies to address these as listed in Section 2.5 on page 55. However, since the intention of this thesis is to build upon the work of the CTML the key principles that will be explored in relation to timelines are repeated below, this is followed by clear definitions of the terms used in this thesis, as derived from the discussion of the literature:

**Multimedia principle** Learning is increased when the materials used contain words and pictures rather than just words.

**Split Attention** Learning is increased when learners do not have to split their attention between several sources of information.

**Signalling/Cueing** Learning is increased when key information is pointed out by visual cues such as colour coding or bold text.

**Generative Drawing** Encouraging learners to produce their own diagrams while learning can increase learning.

## 2.4 Definitions

**Learning:** In this thesis the definition of learning is based on Paas and Sweller (2014, p. 41), who described learning as "Any change in long-term memory involving an accumulation of information". In this work, learning is subdivided into comprehension (initial understanding), and retention (recalling the information from memory). While learning is usually considered to have a further level whereby the knowledge can be reapplied to a new concept, sometimes referred to as transfer, this is typically a level that is reached at a later, more "expert" stage. This aspect of learning is not being considered in this thesis, the work is exploring the essential primary steps of comprehension and retention which are vital as a foundation for the transfer stage. The information that is acquired by participants in the studies described in this thesis is generally either of a historical or temporal nature itself, or is more general information associated with a specific point on a time line or within a corresponding piece of text.

**Comprehension:** The ability to extract relevant information and answer specific questions with study materials available.

**Retention:** The persistence of the information extracted from study materials through comprehension. Studies in this thesis will explore immediate retention (ability to answer the questions without the material present) and delayed retention (ability to answer the questions, without the materials and after a delay, typically a week) this will help to identify

whether the information from the material has been learned and will be available in the future, e.g. for examinations.

**Timeline:** Based upon the Oxford English Dictionary (2012) definition, “a graphical representation of a period of time, on which important events are marked”, timelines in these studies refer to two dimensional depictions of historical information in a graphical format.

## 2.5 Aims and Overall Hypotheses

This research seeks to identify any significant comprehension and learning differences between chronological information presented in text form and presented graphically in the form of timelines. It seeks, in general, to measure both initial comprehension/understanding and the level of retention.

- Are timelines as effective as politicians and education advisors suggest, in particular is information presented in a timeline format more easily understood than text when people are learning?
- Furthermore, are timelines effective in promoting both comprehension and retention aspects of learning?
- Since timelines can be considered an example of multimedia combining both graphics and text, can a commonly used theory of multimedia learning - Mayer’s Cognitive Theory of Multimedia Learning (CTML) be extended beyond its most frequent application in the fields of science, technology, engineering, and mathematics (STEM) to apply to timelines as well?
- When people are studying materials for learning, do their response times for comprehension questions correlate with self reported measures of cognitive load?

These aims are refined into the five hypotheses covered by the studies described here :-



**Hypothesis 1: Use of timelines will result in increased comprehension of information when compared to text.**

**Hypothesis 2: Use of timelines will result in shorter question response times during comprehension of information when compared to text.**

**Hypothesis 3: Use of timelines will result in greater retention of information when compared to text.**

**Hypothesis 4: Individual creation of timelines will result in greater retention of information when compared to provided timelines.**

**Hypothesis 5: Question response times will have a positive correlation with self reported estimates of cognitive load.**

## **2.6 Relationship between Studies and Hypotheses**

Each of the hypotheses is addressed by at least one of the studies as shown in Table 2.2. In terms of age range investigated, study one, being school based, was concerned with students ranging in age from 11 to 14, the other three studies involved adult participants.

## **2.7 Summary**

This chapter has provided a background to the main theories used in the design and analysis of the studies involved in this thesis. Some of the theories were originally developed some

Hypothesis	Study one	Study two	Study three	Study four
Use of timelines will result in increased comprehension of information when compared to text.	✓	✓	✓	
Use of timelines will result in shorter question response times during comprehension of information when compared to text.		✓	✓	
Use of timelines will result in greater retention of information when compared to text.		✓	✓	✓
Individual creation of timelines will result in greater retention of information when compared to provided timelines.				✓
Question response times will have a positive correlation with self reported estimates of cognitive load.		✓		

**Table 2.2** – Hypotheses and their Associated Studies

time ago and updates and adaptations have been indicated where relevant. The theories described are used in a number of the studies in this thesis and the most pertinent aspects are discussed in more detail within the individual study chapters. In the following chapters of this thesis four distinct studies are described. Each includes a rationale, description of the study and a discussion of the results. The studies begin with a real world investigation of timeline use in schools, which is used, to both set the context, and guide requirements, for later laboratory based studies.

# The Battle of Lemonton

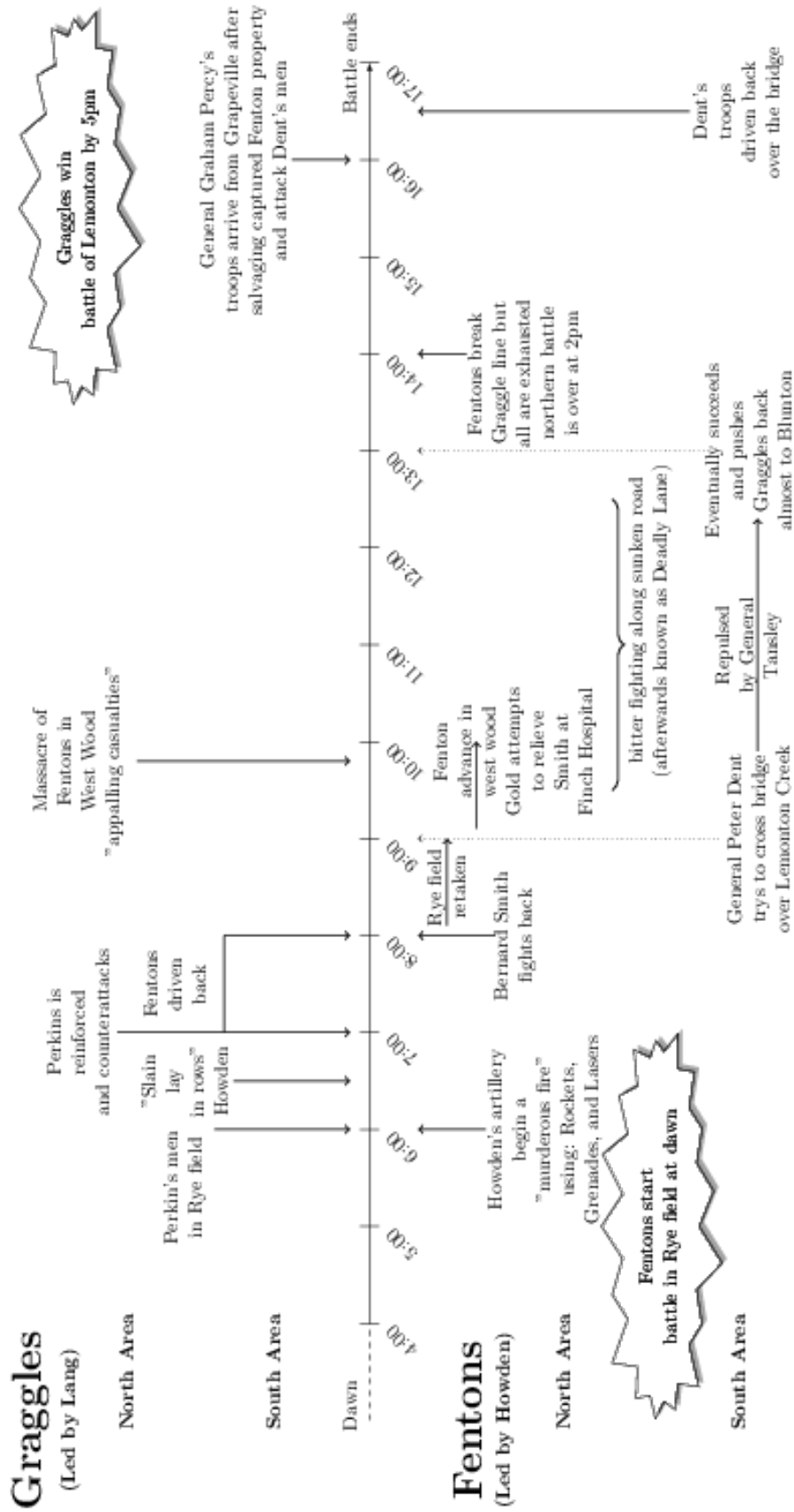


Figure 2.2 – Example of a simple timeline

## **Chapter 3**

# **Study One: A study of using a timeline in a UK secondary school**

### **3.1 Chapter Overview**

This chapter describes a study performed in a secondary school, examining the comprehension benefits of using timelines in the teaching of history. Ecological validity was maintained throughout by using authentic teaching materials, delivered by class teachers, who also performed the assessment. History was selected as the subject since the use of timelines in the teaching of history is promoted by many involved in the provision of primary and secondary education (Hodkinson, 2011; Maddison, 2011). The study found that the mean improvement score for a group of students who had been provided with a timeline was higher than that for a similar group using text alone, and that the difference was significant.

## 3.2 Introduction

### 3.2.1 Rationale

More than most other disciplines, the teaching of history within a national education system has been the focus of a great deal of attention from politicians as well as practitioners and researchers in education (Cruse, 2011; Keating & Sheldon, 2011). There is often contention revolving around not only what should be taught (which events to exclude, which to include, and from which perspective), but also how it should be taught, particularly the balance between memorising basic "facts" and the learning of analytical skills to enable subsequent discovery of further information. The Department for Education (DfE) in England see the development of an understanding of chronology as a key requirement for history teaching at both primary and secondary school level, stating, in respect of students, that "They should know where the people and events they study fit within a chronological framework" (Department for Education, 2013b, p.2), and that they should "know and understand the history of these islands as a coherent, chronological narrative, from the earliest times to the present day. . . ." (Department for Education, 2013a, 2013b). However Maddison (2014), the national lead for history at Ofsted<sup>1</sup>, when summarising lessons learned from school inspections, criticised the level of chronological understanding of many pupils, describing it as "not good" with historical knowledge being "episodic" (p. 5). It should be noted that the term "chronology" is being used here in a wider sense than common dictionary definitions which usually simply refer to arranging dates/events into an ordered sequence. As observed by Stow and Haydn (2000b), within history education, the term generally encompasses "chronological narratives" and "chronological frameworks", understanding of "historical time", the interconnection between events, dating systems, and also time related vocabulary. This framework and understanding is often referred to as providing a "Big Picture" of history which Ofsted (2007, 2011) link to understanding of

---

<sup>1</sup>The UK Government Office for Standards in Education, Children's Services and Skills

Quotation	Source
"This teaching should be supported by the consistent employment of timelines"	Hodkinson, 2011, p. 6
"Timelines though should not be confined to the introductory lesson of a topic. They should be employed within every lesson"	Hodkinson, 2011, p. 6
"much can be achieved now in improving pupils' sense of time by teachers regularly using timelines"	Maddison, 2011, p. 9
"Timelines are basic tools for developing knowledge and understandings about chronology"	Pickford, 2011, p. 25

**Table 3.1** – Quotations From the Primary Education Journal of the Historical Association Promoting Timeline Usage

wider implications of events, and Dawson (2011) considers to be an important facet of chronological understanding. The metaphoric notion of a big picture can be made concrete in the shape of a visual “map of time” in the form of a timeline (Davis, Bevan, & Kudikov, 2013; Alleman & Brophy, 2003; Korallo, 2010b) and the use of timelines in teaching has a long history, Keating and Sheldon (2011), provide examples from UK government guidance documents in 1928 promoting their use.

Educationalists of all types, including academics, practitioners, and politicians, have promoted the use of time-lines when studying history. Table 3.1 lists some examples taken from a special timeline edition of the Primary Education Journal of the Historical Association. The quotation from Maddison is of particular note due to his position as the Ofsted national Adviser for history at that time. Other writers, such as Brophy and Alleman (2003), Dawson (2011), Masterman and Rogers (2002), and Simsek (2007), discuss uses of timelines in wider education. Concern has been raised, however, by a number of academics (Thornton & Vukelich, 1988; Burny et al., 2009; Stow & Haydn, 2000b), that there is little empirical evidence to support their use. This study sets out to measure the effect of timeline use, as an element of the sequencing and visual representational aspects of chronology (Stow & Haydn, 2000a), on children’s comprehension, in three example topics within a secondary school.

## **Developmental considerations**

The experiments in this study involved the use of timelines in each of the three year groups in Key Stage three (KS3), ages 11-14, and account needs to be taken of the likely range of development in understanding concepts in history, and timeline representations in particular. Piaget and Vygotsky are considered by many to be amongst the most influential of development psychologists, both with a long lasting influence, even though they developed their theories distantly and in parallel. Lourenco (2012) describes them as the “two main geniuses”, and many have examined the similarities and differences between the two (Beilin, 1992; DeVries, 2000; Gillen, 2000; Shayer, 2003). Piaget suggested that children learn first to place events in order, then to understand time intervals, before finally grasping the concepts of simultaneity and the arithmetic of time intervals (Lippitz, 1944; Cooper, 1994). Vygotsky built upon the same major foundations (as can be seen via their citations). Where Piaget focused upon the development of an individual, mainly driven from exploration of the physical world, and measured by the way in which they could explain concepts in language, Vygotsky placed much more weight upon interactions with others, and with the environment. Hoodless (1996) comments that, for Vygotsky, the lack of verbal skill need not imply that a concept is not understood.

A significant result of Piaget’s work, and that of many others who subsequently built upon his work, was a widely held view that history could not be successfully taught until adolescence (Hoodless, 1996). An influential report in England "Children and their Primary Schools" (CACE, 1968) (usually referred to as "The Plowden Report" after Lady Plowden who chaired the committee involved) based many of its proposals on this view and had a lasting influence upon history teaching (Thomas, 1990). A later study in the UK (Hallam, 1972) reinforced this, and the view held sway until the late 1980s (Hoodless, 1996).

More recently however, studies by historians, educators, and psychologists (Friedman, 1982; Dawson, 2004; Hoodless, 1996; Stow & Haydn, 2000b) have explored children’s understanding of time in a variety of ways and have often differed from the views of Piaget.



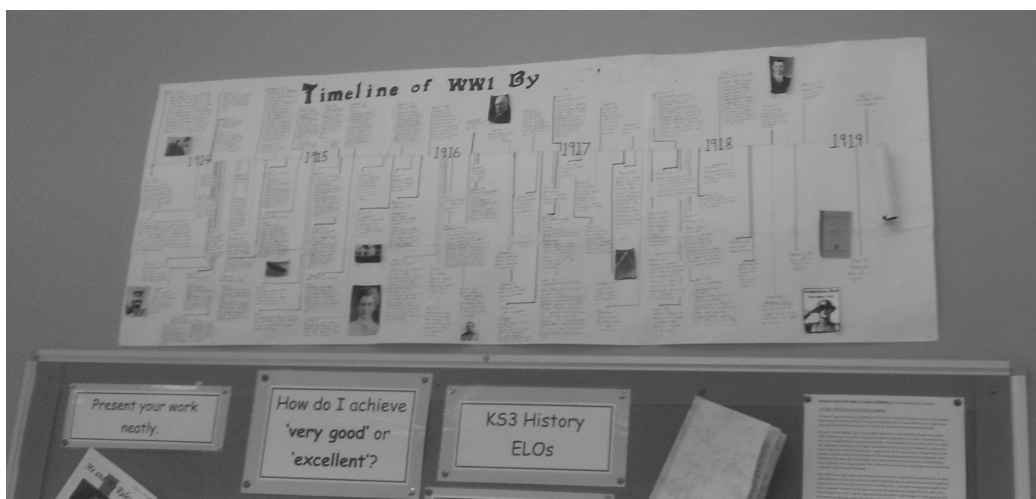


**Figure 3.1** – *Example of a classroom timeline*

Hoodless (2002) points to today's children having much greater exposure to travel, media, IT, etc., and hence a much better understanding of time. Her research with children aged between 3 and 9 years and using conversations about stories as the material for the analysis, showed that, in the right context, use of temporal terms and ability to sequence chronologically is often present at these ages. Burny et al. (2009) quoted from a variety of studies showing significantly greater understanding of temporal matters at much earlier ages than Piaget would have suggested. The teachers taking part in the development and delivery of the present study were confident that the students involved would readily adopt the timelines involved, having been used to using timelines in their primary education and being surrounded by their use in the history classroom, pointing out the examples on the wall shown in Figures 3.1 and 3.2 on the next page.

The study maximised ecological validity through use of topics from the existing curriculum, delivered in a standard classroom setting by the usual teachers of the students, as recommended by Black and Wiliam (1998).

In terms of assessment of the value of timelines in studies, to maintain ecological soundness, the study uses the school's own assessments. For many years, learning was simply assessed by using individual scores in tests and examinations, with a student's achievement being compared against that of peers in the same school or nationally, with each of these snapshots



**Figure 3.2** – *Example of a student produced timeline*

considered in isolation. While this approach remains in place for many of the nationally set examinations (most GCSE examinations in the UK for example), more sophisticated approaches are now used within schools, and by organisations such as DfE looking at the performance of schools as a whole. These approaches recognise that learning is a continual process involving progress/growth as well as attainment. Anderman, Gimbert, O’Connell, and Riegel (2015) highlight the complexity of examining and understanding change in educational environments, and describe a range of approaches, based around the notion of a path or "learning trajectory" through different stages and levels. Research by the DfE (2011) uses the term "learning pathways" rather than trajectories and recognises that these are not inevitably monotonically increasing, since some students may regress to earlier attainment levels or stay static for a period, and that progress also may not be linear, with more progress being made in KS2 than KS3 and more during summer terms than in winter<sup>2</sup>. Since the school was able to provide the target level and the previous assessment level for all of the students involved in the study, these were combined in an approach (described in 3.3.4 on page 71) that made full use of all of the data as trajectories.

<sup>2</sup>Although the report DfE (2011) relates to progress in reading, writing, and mathematics, the broad consistency of results suggests that the same is likely to apply to history

### **3.2.2 Aims**

This study has been developed to examine quantitatively some practical examples of the use of timelines in an English secondary school setting, prioritising ecological validity. The hypothesis considered in this study (numbered as in section 2.5 on page 55) is:-

Hypothesis 1: Use of timelines will result in increased comprehension of information when compared to text.

## **3.3 Method**

### **3.3.1 Participants and Recruitment**

The study was granted ethical approval from the Department of Psychology ethics committee (ethics reference number:15-255). The school used for the study was a comprehensive, mixed gender, secondary school (covering age range 11-18), in south west England. The school roll was around 1300 students with an intake of around 230 students each year. To reduce the jump in scale that might otherwise be experienced by students arriving from small feeder primary schools, and to ease the provision of pastoral care through smaller numbers in tutor and year groups, the lower school (Key Stage three (KS3), ages 11-14) was managed and run by being split into two separate ‘populations’. The school balanced these populations in respect of gender and abilities (based upon input data from the primary schools) through allocation on entry to the school, and the split applied to teaching as well as pastoral care. The students within each population were further divided, according to ability for particular subjects, or groups of related subjects (e.g. humanities), into sets (upper set - set one, middle set - set two, and sets three and four which were both mixed lower ability), with resultant set sizes of around 30. Key stage three contained three year groups, years 7, 8, and 9, (Y7, Y8, Y9), and so, with four sets per year group, in each of the two populations, there was a total of 24 sets for each subject. In the case of history these

were taught by a group of four teachers with each set having an allocated teacher. The study involved making an additional use of timeline material in one set of each year group and comparing their results with those of the corresponding control set in the same year group. The two teachers who agreed to collaborate in the research each had responsibilities for a range of sets but a common factor was that they both taught a set one group and so it was agreed that set one groups would be used for the experiment. One of the teachers involved in the study left the school part way through the study and this prevented the analysis of the year 7 (age 11-12 yrs) group since the relevant assessment data had not been collected at the time of leaving (the approach and materials used are still described later for completeness however). For the remaining year 8 and year 9 groups, a total of 102 students took part in the study, 49 in year 8 (age 12-13 yrs), and 53 in year 9 (age 13-14 yrs). Of these, 50 were female and 52 male. The detailed group breakdown is shown in Table 3.2.

	Female	Male	Total
Y8 control	12	13	25
Y8 Timeline	9	15	24
Y9 Control	14	15	29
Y9 Timeline	15	9	24

**Table 3.2** – Study One - Numbers in Each Group

### 3.3.2 Materials and Measures

The school operated a two week timetable, with history being taught for one lesson per week in year 7, and three lessons per fortnight in years 8 and 9, leading to a maximum of 39 lessons in a year for year 7 and 56 for years 8 and 9 (the figures are maxima since in any one year public holidays, and revision periods, can reduce these totals). The overall total of around 150 lesson periods were used to cover the broad themes for the history KS3 curriculum, which were set nationally (Department for Education, 2013a). To identify

appropriate experimental materials, an examination was performed of all the lesson resources used by the school for KS3 history (reviewing presentations, notes on expected learning outcomes, and assessment materials), categorising each lesson as to whether it already used timelines in some form (18 instances), whether the use could be further extended (seven instances) or whether a timeline could be meaningfully incorporated into a lesson that did not currently use them (11 instances). This analysis was discussed with the teachers involved in the study, examining those lessons where new or additional use of timelines could be made, while taking account of their teaching workloads and external constraints (holidays, examinations, etc). The option chosen, for each of the year groups, was to use a completely new timeline for a topic being studied during the penultimate term of the school year (term five). The researcher then searched for, and provided, suitable timeline material (ensuring that these were suitable in content, copyright etc.). For year 7 the topic was "Religious changes in Tudor times" (see Figure B.1 on page 238), for year 8 the topic considered was "What was the most significant 'moment' in the development of the English language?" (see Figure B.2 on page 239), while year 9 considered, "What was the most significant event of World War Two?" (see Figure B.3 on page 240). Timelines can take many forms, and the three used here reflect some of that diversity. The year 7 timeline used royal reigns to indicate time periods rather than numerical dates and used the vertical dimension to indicate the contemporary extent of religious constraint, with cartoon pictures used to illustrate events and flow. The year 8 timeline continued the use of images, included dates, and used a staggered timeline approach with a non-linear scale. The year 9 timeline was the most text heavy of the three, with no images used, and extending the time flow into the vertical dimension when listing some sets of close events. Table 3.3 summarises the timelines used.

To maintain ecological validity the study did not involve any separate assessment measurements, but instead used the normal school assessment process which allowed the teacher, in consultation with the head of department, and school leadership team, to select

Study	Linear	Time marked	Linked detail	Multiple timelines	Complexity	Images	CTML Principles
1 - Y7	No	Only by reigns	yes panels containing cartoons	No	single line, 4 events	Yes (plus timeline itself)	CTML, Split Attention, Spatial Contiguity, Temporal Contiguity, Coherence, Signalling
1 - Y8	No plus 3 line zig zag	Dates but dates mixed with centuries	Yes	No	12	Yes	CTML, Split Attention, Spatial Contiguity, Temporal Contiguity, Coherence, Signalling
1 - Y9	Yes horizontally but not vertically	Years horizontally, specific months vertically	Yes (vertical lists)	vertical detail	8 primary dates, 19 in verticals	No	CTML, Signalling

**Table 3.3** – Summary of Characteristics of Timelines used in Study One

the most appropriate method for the topic. Student progress assessment took place at the school, across all subjects every term and used a variety of methods. In some terms specific written work produced during the term was assessed, in some others a teacher rating (after considering sample pieces of work) was used, while in further cases a written one hour formal exam could be used. For this study, the Year 7 group students were required to produce a significant piece of homework over the term (known as an ELO, Extended Learning Opportunity), while the Year 8 and 9 students would produce essays on their respective topics. Each teacher would mark their sets as usual and consistency of marking, between the modified and control sets, was provided via discussion between teachers involved and by the regular quality review by members of the school's leadership team (both activities being a normal part of the department/school quality regime).

### **3.3.3 Design**

A 'between group' design was used. Since the lower school (KS3) was already split into two roughly equal populations, one set in each year group would be taught a topic using a timeline in addition to the normal textual materials, while the corresponding set in the other population would only use the textual materials. As described in Section 3.3.1, two teachers agreed to take part and the sets that they taught (which were all set one, higher ability) would use the timeline material. The two teachers would incorporate this material into their normal lesson and assessment. The independent experimental variable for each student was therefore their membership of the timeline or the control group. The dependent variable was the difference between the student's mark resulting from the experiment and the mark that they would have been expected to get from linear interpolation between their mark for the previous term and their target mark for the final term (the derivation of this variable is described in more detail in Section 3.3.4).

Sub-Level	Meaning
C	About to begin to learn the more complex areas of a new level.
B	Have a good, solid, understanding of the components within that level
A	Have a full understanding of all the topics within that level and are now ready to learn the more challenging areas

**Table 3.4** – National Curriculum Sub Levels

### **3.3.4 Procedure**

In line with the requirements of the school that the study be performed as an evaluation of changes of the sort normally made by teachers when wishing to trial a new piece of material, and to ensure that the study was as authentic as possible, the incorporation and delivery of material was performed by the teacher during a normal lesson (no researcher present). The students were not made aware of the study until after its completion. The collected data were anonymised using unique student numbers allocated by the school for the duration of the study linking academic and gender data.

### **Data Preparation**

The school used the standard rating system specified within the UK KS3 national curriculum at that time, with levels that increase from three to seven and sub-levels A, B, and C within each level. The sub-levels reflect developmental stages as shown in Table 3.4 and the levels themselves are also relatively subjective, at one level a student will be considered able to describe events, in the next level to explain them and, in the next, to also explain causes and consequences.

Although ordinal, therefore, neither the levels, nor the sub-levels, correspond to equal intervals. For educational analysis, however, it is usual to treat them as an interval scale (Blatchford, Russell, Bassett, Brown, & Martin, 2007; Borra, Iacovou, & Sevilla, 2012; Farrell, Dyson, Polat, Hutcheson, & Gallannaugh, 2007) by using a conversion such



Grade	Average Point Score
4C	25
4B	27
4A	29
5C	31
5B	33
5A	35
6C	37
6B	39
6A	41
7C	43
7B	45
7A	47
8C	49

**Table 3.5** – Conversion of Grades to Average Point Scores for Analysis

as that shown in Table 3.5, and the academic data from the school were converted this way. For each student the school provided their gender and a total of three data points, their score for the topic in the study, their score for the previous term (term four), and the target level that the teacher had set at the beginning of the school year, which they would therefore be expected to achieve by the end of the following term (term six). A possible analysis approach could have been to simply compare the results for the previous term with those for the current term using a two way ANOVA. However Willett (1988) highlights the limitations of approaches in educational research that are based upon simple pre and post tests, and encourages instead, the use of multiple measurements and the production of individual trajectories for each student. The availability of only three data points per student precluded the use of the sophisticated analysis approaches advocated by Willett (1988) and also those described by Anderman et al. (2015), but allowed the term four and term six values to be used to define a two wave linear trajectory and hence provide a better baseline from which to

measure the effect of the intervention. The dependent variable was therefore defined as the difference between the mark obtained from the experimental assessment and the mark that could be expected by linear interpolation between the term four mark and the term six target. The sets of marks were examined visually, and by calculation, for normality of data, skewness, kurtosis, and variance. Appendix F contains the results in Tables F.1, F.3, and F.2. In the majority of cases the assumptions of normality were not met, according to the Shapiro-Wilk tests. The skews were both positive and negative and so there was no suitable transform that could be uniformly applied (Field, Miles, & Field, 2012). For the calculated mark difference, used as the dependent variable, the results of the corresponding analyses are shown in Table F.4 where it can be seen that all but the Y9 Timeline condition can be considered normal at a probability level of ( $p > .05$ ). However, examining each distribution for outliers, using a threshold of three times the absolute deviation around the median (Leys, Ley, Klein, Bernard, & Licata, 2013), showed that there were a number of outliers in every case.

The first step before analysing the results of the intervention was to examine the equivalence of the timeline and control groups for each year group in terms of previous and target marks. Since the sampling distributions could not be considered normal, the use of t-tests was discounted (Field et al., 2012). The non-parametric tests for independent groups most often considered equivalent, the Mann-Whitney test and Wilcoxon's rank-sum test (Field et al., 2012), are also considered by many statisticians to provide poor performance, especially regarding type I error rates, in the presence of differences in sample sizes, distribution variances, and shapes, all of which applied to the marks in this data set (Wilcox, 2016; Fagerland & Sandvik, 2009; Keselman, Algina, Lix, Wilcox, & Deering, 2008). Robust testing techniques, as recommended by Wilcox (2016) and Kang, Harring, and Li (2015) were therefore used for all the analysis steps. Although the gender of each student had been provided, groups split by gender were too small to allow statistically meaningful comparisons to be made (Luh & Guo, 2007).

## 3.4 Results

### 3.4.1 Year seven results

Although the experimental intervention took place, the departure of the teacher involved meant that no results were available for this group.

### 3.4.2 Year eight results

#### 3.4.2.1 Group comparisons of year eight data

Before analysing the results from the study, the two groups were examined for balance. The groups did not differ significantly in terms of their proportion of gender,  $\chi^2(1) = .206$ ,  $p = .65$ , and no significant differences in 20% trimmed means between the control group ( $M = 31.9$ ) and the timeline group ( $M = 32.5$ ) were found for the previous (term four) result when assessed using the Yuen-Welch method (Wilcox, 2016),  $p = .58$ , 95%  $CI[-1.53, 2.66]$ , nor for the 20% trimmed means of the target scores (term six), control group ( $M = 36.7$ ), timeline group ( $M = 36$ ),  $p = .325$ , 95%  $CI[-2.24, 0.78]$ . The two groups were therefore considered to be suitably balanced.

#### 3.4.2.2 Analysis of year eight results

The marks difference from expected values for the Y8 timeline group with a 20% trimmed mean ( $M = 0.69$ ), showed greater improvement than the Y8 control group, with a 20% trimmed mean ( $M = -1.1$ ) when assessed using the Yuen-Welch method,  $p = .005$ , 95%  $CI[0.59, 2.9]$  with a large explanatory measure of effect size ( $\hat{\xi} = 0.5$ ), roughly corresponding to a Cohen's  $d$  value of 0.8 (Wilcox, 2016; Ellis, 2010).

### **3.4.3 Year nine results**

#### **3.4.3.1 Group comparisons of year nine data**

Before analysing the results from the study the two groups were examined for balance. The groups did not differ significantly in terms of their proportion of gender,  $\chi^2(1) = .575$ ,  $p = .44$ , however significant differences in 20% trimmed means between the control group ( $M = 37.8$ ) and the timeline group ( $M = 35.5$ ) were found for the previous (term four) result, when assessed using the Yuen-Welch method,  $p < .001$ , 95%  $CI[1.37, 3.07]$ , for the 20% trimmed means of the target scores (term six), control group ( $M = 41.25$ ), timeline group ( $M = 39.63$ ),  $p = .015$ , 95%  $CI[0.34, 2.89]$ , and also for the expected growth (as shown by the difference between the term four mark and the term six target), where the 20% trimmed means for the control group ( $M = 4.35$ ), and timeline group ( $M = 3.58$ ), differed significantly,  $p = .04$ , 95%  $CI[0.06, 2.53]$ . The two groups were therefore considered not to be sufficiently balanced for direct comparison.

#### **3.4.3.2 Analysis of year nine results**

Although, as discussed above, the significant differences between the two groups in respect of target, previous, and expected, scores suggest that direct comparisons of results should be treated with caution, the comparison was performed in order to inform the discussion below. The distribution of marks differences from expected values for the Y9 timeline group with a 20% trimmed mean ( $M = 0.79$ ), did not differ significantly from the Y9 control group, with a 20% trimmed mean ( $M = 0.86$ ) when assessed using the Yuen-Welch method,  $p = .597$ , 95%  $CI[-2.25, 1.32]$  with an explanatory measure of effect size ( $\hat{\xi} = 0.12$ ), roughly corresponding to a Cohen's  $d$  value below 0.2 (Wilcox, 2016; Ellis, 2010). A post hoc exact grouping analysis was also performed. For this analysis two groups (Timeline and Control) were formed and participants were paired between these so that each participant in a group was paired with a single participant with identical target and previous

scores in the other group. This analysis is provided in Appendix G but also did not support the hypothesis in respect of Y9 results.

### **3.5 Discussion**

Studies investigating the principles of Mayer's Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2005) have shown that using learning materials that obey these principles can increase the comprehension and retention of information by students (Massa & Mayer, 2006; DeLeeuw & Mayer, 2008; Mayer, 2009a). Comprehension and retention are fundamental aspects of learning and so educators and researchers are keen to develop effective materials and processes to assist. Most of the existing CTML research has focused, so far, on science based material, and the majority of the experiments have been laboratory based, and relatively brief (Ayres, 2015). Meanwhile, for teaching the subject of history in the UK, there has often been strong promotion for the use of graphical timelines (Keeting, 1976; Dawson, 2011), which, as they generally combine both graphics and text, are good examples of multimedia outside of the field of science education. Some researchers, however, have questioned this strong promotion, citing a lack of evidence of effectiveness (Thornton & Vukelich, 1988; Burny et al., 2009; Stow & Haydn, 2000a). This first study sought to explore timelines as an example of multimedia situated in the real world setting of a school. The opportunity to perform a study in a working school, at an early stage, was welcomed, as it would set an ecologically sound reference point for other, laboratory based, studies.

In ecological terms, research in schools involves striking a balance between the internal and external validity of the experiment, with the former usually requiring highly controlled, but artificial, conditions, and the latter generally involving a more natural, and less interventionist, approach (Slavin, 2008; Black & Wiliam, 1998; Ross, Morrison, & Lowther, 2010; Ross & Morrison, 1989). The study was designed to focus on hypothesis one: "Use of

timelines will result in increased comprehension of information when compared to text” and it was expected that the links between time and space in language (Casasanto & Boroditsky, 2008; Vallesi et al., 2008), together with the associated concept of a Mental Timeline (Bonato et al., 2012), would make a timeline representation a natural way to convey chronological information. This, together with the use of multimedia principles from the CTML to guide its presentation (Mayer, 2014a), would lead to measurably greater comprehension of the underlying information by students. The study was conducted with 3 year groups (although data was not available for analysis for one of these, the Y7 group). Analysis of the Y8 results showed that the group using the timeline achieved significantly higher results against expected values when compared to those of the control group with no timeline material. This result is in line with hypothesis one, and with the underlying multimedia principle that learning is increased when the materials used contain words and pictures rather than just words. Analysis of the Y9 results (where the group baselines were not balanced, but where interpolation and sub-setting approaches were taken during analysis in an attempt to ameliorate this), however, did not support the hypothesis. To discuss these mixed results it is necessary to explore the design compromises made in order to maximise ecological soundness, which may well have led to inherent variability in the results, primarily due to the limitations/inflexibilities regarding choice of assessment method, choice of material, consistency of assessment, consistency/transparency of delivery, and availability of data (both primary and secondary).

### **Assessment method**

The assessment method had already been set by the teachers involved, as part of their plan for the year. This took the form of a marked essay, created as a homework exercise, with all materials available to the student. To produce these essays would have involved the students marshalling arguments (McCutchen, 2000) derived from the various sources (Seufert, 2003). Although there would be an element of short term retention involved regarding which facts

to include in the essay and where to find them in the materials, it was primarily comprehension that was being assessed (McCutchen, 2000; Mason, Davison, Hammer, Miller, & Glutting, 2013). The timeline materials should, therefore, be considered in respect of the ease of accessing the information that they were intended to convey (Canham & Hegarty, 2010; Eilam & Ben-Peretz, 2010).

### **Choice of materials**

The additional timeline materials had been selected by the teachers, following discussion with the researcher, to be compatible with the baseline materials (text books, handouts, and class work) that would be used across both sets in each year group. Table 3.3 summarises the characteristics of the three timelines used. The timelines share some important similarities; each can be considered, as an entity, to uphold the Multimedia (Mayer, 2014a; Butcher, 2014) and Signalling (van Gog, 2014; Richter, Scheiter, & Eitel, 2016) principles of the CTML, but, while the Y7 and Y8 timelines also obey the Split Attention (Florax & Ploetzner, 2010), Spatial Contiguity (Mammarella, Fairfield, & Di Domenico, 2013), Temporal Contiguity (Mammarella et al., 2013), Coherence (Mayer, 2014a), and Redundancy (Mayer & Johnson, 2008) Principles, the Y9 timeline mostly consists of text. It only contains a single highlighted line, plus its overall structure in terms of blocks of text, that could be considered as images, and so it cannot be said that these additional principles are relevant/met in the case of the Y9 material and this may explain the lack of effect seen in the results.

Since the experimental groups received the timelines in addition to the common text materials it may also be that the combination of timeline and textual materials in the Y9 case resulted in a lesser degree of Spatial Contiguity, and greater Split Attention involved when the total set of materials (text and timeline) are considered. Therefore, the additional cognitive load involved in integrating material from multiple sources (Westelinck, Valcke, De Craene, & Kirschner, 2005; Ainsworth, 1999; Wineburg, 1991) may have outweighed

any timeline benefits and hence did not result in support for hypothesis one in the case of the results for Y9.

The timelines varied between years although each has a primary flow from left to right (earlier events on the left), in line with much of the research that relates the expected direction of mental timelines in western culture to the direction of reading (Pitt & Casasanto, 2016). However, in the case of the Y8 timeline (Figure B.2), that linkage with reading is taken even further as, rather than being a single line, the timeline zig zags down the page requiring gaze shifts similar to those expected when reading a page of text (Liversedge et al., 2016; Rayner, 2015). The Y9 timeline (Figure B.3) has a primary flow of years from left to right, but extends this, with lists of events that occurred in the particular year being shown vertically (earliest at the top). Furthermore, the Y7 timeline (Figure B.1), whilst simple in temporal terms, with only four events marked, uses the vertical dimension to indicate the extent to which the prevailing environment was authoritarian vs liberal (by the height of the roller coaster). It could therefore equally well be considered as a graph of authoritarianism vs time. The Y9 timeline is completely text based, while the Y7 and Y8 timelines incorporate images in the form of cartoons or drawings to convey aspects of the information and hence, had this been a memory retention based study, their results may have benefited more from the multimedia effect (Mayer, 2014a), and from the picture superiority effect (Shepard, 1967; Paivio & Csapo, 1973; Whitehouse, Maybery, & Durkin, 2006; Paivio, Rogers, & Smythe, 1968). Furthermore, as suggested earlier, even the short term memory retention involved in marshalling arguments for an essay may have benefited from the incorporation of these images, contributing to the significant effect in the case of Y8 and to the absence of any significant effect in the Y9 case.

None of the three timelines can be given a simple, single categorisation as an image in this way however. The Y7 timeline is a composite diagram with four sub-panels, each with three or four cartoons and also cartoon characters in the main timeline. The Y8 timeline similarly incorporates 12 sketch pictures plus diagrammatic arrows. The Y9 timeline did not use



images, being purely text based (apart from the primary timeline being highlighted as a horizontal shaded bar), but its overall structure was relatively dense and the student reader's attention might be expected to wander around the page without the clear signposting that additional images might provide. The horizontal timeline increases from left to right, while the vertical sub-timelines increase in date from top to bottom, but this could also be seen as inconsistent in terms of earliest sub-line dates being closest to the main line below the line, but furthest away when the sub-line is in the top half of the page. It would be instructive to investigate all three timelines with the benefit of eye-tracking equipment Mayer (2010), Jamet (2014), Rey (2014), Lindner, Eitel, Strobel, and Koller (2017), Johnson and Mayer (2012). The timelines (certainly in the graphic heavy Y7 and Y8 examples) are best considered not only as multimedia in conjunction with the common text provided to both experimental and control groups, but also as multimedia in their own right combining text, diagrams, and structure.

For this study, the materials were selected in conjunction with the teachers involved, rather than being designed to explicitly uphold the principles, but their inherent compatibility and positive result appears to support, or at least not contradict, those principles in the case of one year group.

### **Variability of assessment**

In terms of scoring the assessment material, the approach was defined by the school and used the common approach of each teacher marking their students' work but using discussion amongst peers and review by senior leaders to level the scores and provide consistency. Again, the use of existing practice (also common in other schools) brought an advantage in terms of external validity, and a clear consistency with the approach that would have been used for the previous assessment and also in setting target scores (both of which were important data points for the analysis), but it could be criticised in terms of the experiment, regarding imperfect objectivity and lack of a blind marking approach. Also, since neither

the intervention, nor the control, lessons were observable by the researcher, these may have been different in some way in style of presentation, or mood of the class. Various class observations undertaken by the researcher as a school governor have highlighted the effect that the mood/atmosphere of a lesson can have, at least on the perceived learning environment, and it could be that this may explain some of the variation in results.

### **Variability of groups**

Perhaps the most important compromise made between internal and external validity, however, concerns the inability to assign participants to experimental groups (whether randomly or against criteria such as existing scores/targets). Slavin (2008) suggests that this is one of the most contentious aspects of syntheses of education program evaluations, considering it essential to use random group assignment to meet the highest standards, but acknowledging a degree of trade off with group size. Meanwhile Ross et al. (2010) and Black and Wiliam (1998) prioritise external validity, as was done in this study. The experiment made use of an existing split in the school population to define the groups and, whilst the entry process at Y6/7 aims to balance those in terms of ability, the Y9 groups in this study were found to be significantly different in respect of both their target scores (reflecting teacher judgments of abilities), and their scores on the previous assessment (reflecting abilities in that task). These differences could also have been compounded by effects due to levels of prior knowledge on learning from both the text materials and the added multimedia timelines (Kendeou & Van Den Broek, 2007; Kalyuga, 2013). Also, since assessments were produced as homework, there was no control over additional material and it may be that some of the participants in the control group obtained and used other graphical timeline material. A number of approaches were taken to post-hoc matching of subsets of the groups as well as using the planned interpolation approach, however, none of these analyses identified significant differences in distributions that would support hypothesis one. The groups in Y8, however, did not differ significantly on distributions of

scores for their targets nor for their previous assessments and hence were considered to provide a meaningful baseline and better founded support for hypothesis one.

### **Availability of data**

The school was only able to provide three assessment data points for each student involved, their score for the intervention assessment, their previous score, and their end of year target. This was enough to perform an interpolation, trajectory, based analysis, but additional data points could have produced more robust results (Willett, 1988; Anderman et al., 2015), and it is also unfortunate that results were only available for two out of the three year groups.

### **Summary and next steps**

In summary, this study provides some support for hypothesis one, but many of the compromises made for ecological validity (in particular complexity of materials, potentially biased delivery/marketing, and unbalanced grouping) mean that the result should be considered illustrative of potential rather than conclusive.

Other research in the area of multimedia learning suggests that a number of factors such as spatial ability (Hoffler, 2010), gender (Wilkening & Fabrikant, 2011), and learning preferences (Mayer & Massa, 2003; Wilkening & Fabrikant, 2011) may have an effect on learning from diagrams, maps, and other multimedia. It was not possible to collect data regarding spatial ability and learning preferences in this study (due to the need to maintain ecological validity, and the time pressures on teaching) and the set sizes were too small for further sub group analysis based on gender, so no conclusions can be drawn on either aspect. Finally, the groups may not have been equal in their level of prior knowledge of the subject Black and Wiliam (1998) note, in their review of a range of different educational studies, that the closer a study comes to ecological validity, the harder it is to perform a quantitative comparison against a truly equivalent control. It was therefore decided that the next set of studies would be performed in observable conditions, would include the collection of spatial

ability, multimedia learning preferences, and relevant demographic variables and would use fictional material.

This study examined comprehension since the material was always available for the participant to refer to as they produced their work for assessment. If, however, the teachers had chosen to use a closed book style examination based assessment mechanism at the end of the topic, then it would have been the differences in participant retention resulting from the inclusion of timeline material that would have been analysed for the study. Since retention or long term change is an important part of learning (Kimble, 1964; Paas & Sweller, 2014), the next study incorporated assessment of retention as well as comprehension.

In summary, this study has shown some support for the hypothesis that use of timelines will result in increased comprehension of information when compared to text, with the results for one year group, of the two where data was available, differing significantly, in line with the hypothesis. For the other year group with data, it has been observed that the timeline used contained very little graphical structure, being mostly text, and that the resulting inapplicability of many of the CTML principles, may have led to the lack of evidence to support the hypothesis. The study overall had a strong focus upon ecological soundness, and various compromises were necessary to achieve that. The study does however provide a useful starting point and context for the research, which aims for relevance in an education environment. For the next study, the materials would use more traditional timeline forms, the study would be performed in more controlled conditions, supervised by the researcher, would use objective, computer based, testing, and would use fictional materials to remove any potential confounds from prior knowledge.

## **Chapter 4**

# **Study Two: An experimental study on the use of timelines**

### **4.1 Chapter Overview**

The first study, performed in an operational school environment, supported the hypothesis that comprehension of information can be increased by the use of timelines, as suggested by the multimedia principle within the CTML (Butcher, 2014). This, second, study was designed to complement and extend the first by introducing an assessment of retention as well as comprehension, by using a within participant design, in more closely controlled conditions, and by collecting additional demographic and individual difference variables. The study also examined the effects of presentation mode (timeline or text) on the immediate retention of content. During the comprehension phase of the study the material remained visible, while during the subsequent retention phase it was no longer available. The comprehension and retention effects were measured through two dependent variables, participant accuracy, and response time when answering questions. The response time was expected to reflect cognitive load and the relationship between that measure and a self report method previously used by DeLeeuw and Mayer (2008) was also studied. Two secondary

aims were addressed: to ascertain the relationship between participant accuracy and spatial abilities/graphical preferences (as assessed by commonly used test instruments), and to identify any association between participant accuracy and other demographic variables such as age and gender. In contrast to the first study, this study found no evidence that the use of timelines would lead to greater accuracy in comprehension and retention, with the only significant result in the accuracy measures suggesting the opposite to hold in the case of retention. No significant relationships were found between the measurements of accuracy and any of the spatial abilities/graphical preferences, nor with any of the collected demographic variables. There was support for a relationship between self reported cognitive load and response time measurements during comprehension tests, although the self-reported load was not found to be significantly different between the material types. For some comprehension questions however, the response time measurements were found to differ significantly, with responses in the timeline condition being faster. The chapter concludes with a discussion section suggesting that the retention result be explored further by means of a delayed retention test.

## **4.2 Introduction**

### **4.2.1 Rationale**

The studies in this thesis examine various aspects of learning from timelines. The first study explored the use of timelines, in an ecologically valid manner, in an English secondary school history teaching environment, and found an improved learning outcome for year 8 students when timeline material was added to a lesson. However the constraints engendered by maintaining ecological validity for that study raised risks of potential confounds that could not be controlled or measured (levels of prior knowledge, environmental conditions, and subjectivity of marking). This second study was therefore designed to be run under more controlled conditions, to explore both comprehension and immediate retention, and to

examine a range of additional variables that it had not been possible to collect in study one. When considering how well information is initially comprehended, and then retained, the way in which it is presented has been shown to have an effect, particularly when combinations of sensory modes are involved. Common combinations that have been studied are those of text and graphics: (Mayer & Gallini, 1990), graphics and audio: (Mayer & Sims, 1994), as well as less common examples of audio, graphical and enactment: (Dijkstra & Moerman, 2012), and graphical and olfactory: (Lwin et al., 2010). The common principles underlying each of these studies are those of cognitive load, as in Sweller's Cognitive Load Theory (CLT) (Sweller, 1988, 1994, 2010; Sweller & Chandler, 1994; van Merriënboer & Sweller, 2005; Paas et al., 2004), and independent bandwidth limited sensory/processing channels, as in Paivio's Dual Coding Theory (DCT) (Paivio, 1991; Clark & Paivio, 1991). The CLT and DCT underpin a set of more specific theories that are grouped together as the "Cognitive Theory of Multimedia Learning CTML" developed by Mayer (2005), and supported in his many subsequent studies (Mayer, 2009a; Massa & Mayer, 2006; DeLeeuw & Mayer, 2008). The CTML incorporates previous work by Mayer (1996) and Wittrock (1992) concerning active learning and organising/integrating new information with existing knowledge, and has relevance to the study described in this chapter since the timelines incorporate textual information as well as the purely diagrammatic elements.

Where CTML related research has considered diagrams, these are mostly scientific in nature (Mayer, 2005; Mayer, 2009a; Kalyuga, 2008; Mayer, 1989a), where the aim is to help the reader understand an underlying technical concept, rather than the temporal relationships involved in a timeline for example. Providing a few examples outside of science, Larkin and Simon (1987) included graphs, taken from the subjects of economics and physics, in their consideration of types of graphical learning material, while Winn (1991) examined the use of diagrams to convey sequence, which is of direct relevance to timelines as they show sequences of events, including language based sequencing (Winn & Solomon, 1993; Winn, 1988).

This study seeks to assess the relevance of the CTML in the specific area of timelines. The Multimedia Principle itself, that learning is increased when the materials used contain words and pictures rather than just words (Butcher, 2014), forms part of the most basic comparison within the study since timeline materials included both text and graphics while the corresponding text materials only used text. The Split Attention principle - that learning is increased when learners do not have to split their attention between several sources of information, whether physically or temporally split (Sweller, Ayres, & Mayer, 2006; Florax & Ploetzner, 2010; Mayer & Moreno, 1998; Kalyuga et al., 1999), was used in the design of the experiment and materials, where there is only ever a single source shown at a time. The presentation of timelines also follows the Temporal Contiguity Principle - that learning is increased when text and associated graphics are presented together in time (Schuler, Scheiter, Rummer, & Gerjets, 2012; Mammarella et al., 2013) since all of the associated text is included when the timeline is presented. Care was taken during the development of materials to follow the Redundancy principle - that learning is reduced when the same information is provided in multiple forms or is over elaborate with extraneous detail (Morrison, Watson, & Morrison, 2015; Kalyuga et al., 1999; Jamet & Le Bohec, 2007; Mayer & Johnson, 2008; Ari et al., 2014), but there was a fine line to be considered, especially within the text material, between "extraneous detail", and keeping the story engaging. The story was written to be as succinct as possible, containing all relevant facts, but including enough additional descriptive material to provide a flow for the reader. The Signalling/Cueing Principle - that learning is increased when key information is pointed out by visual cues such as colour coding, or bold text (Richter et al., 2016; van Gog, 2014) was only used in the text in so far as the text is structured into distinct paragraphs. For the timelines, however, colour coding was used with battles depicted in red, rulers in blue, culture in green, and general information in black although this was simply an artifact of the timeline generation tool and was not explicitly pointed out to participants. Text labels within the timelines were placed close to the associated event to follow the Coherence Principle



- that learning from multimedia lessons can be increased when the graphics are relevant to the goal, and decreased when they are not (Mayer, 2014a; Harp & Mayer, 1998; Mayer, Heiser, & Lonn, 2001, p. 280), noting, however, that others have not found convincing evidence of the effect (Muller, Lee, & Sharma, 2008; Rop, Verhoeijen, & van Gog, 2017). Prior subject knowledge is known to affect learning for both textual (Ozuru, Dempsey, & McNamara, 2009; Alexander, Kulikowich, & Schulze, 1994; McNamara & Kintsch, 1996; Kintsch, 1994) and diagrammatic material (Vekiri, 2002; Lowe, 1996; Cheng, Lowe, & Scaife, 2001). The potential confound of varying prior subject knowledge can be reduced by the use of fictional materials in experiments. Schnotz and Baadte (2015) took this approach when examining diagrams and Voss and Silfies (1996) did the same for text, although they note that general familiarity with material types (in this case history) did have an effect. There are similarities between the study performed by Voss and Silfies (1996) and this study since they used a matched pair of historical stories, set in fictional countries, used a set of questions as part of the assessment, and used a within participants design. They incorporated an essay question in addition to the multiple choice questions (which this study does not, in order to make all aspects fully objective). However they did not examine retention or response times. The focus of their study was an examination of differences in comprehension between a text and an expanded text contrasting the results for participants with higher levels of comprehension with those with higher levels of familiarity with the types of material.

Assessing the cognitive load experienced during the learning of multimedia material can be done in a variety of ways (Brunken et al., 2010), frequently by using a self assessment scale, the subjective cognitive load measure (SCL) (Paas & Van Merriënboer, 1993; van Gog & Paas, 2008; Paas et al., 2003). Performing the self assessment during the experiment can however be disruptive (Paas & Van Merriënboer, 1994; Chen, Ruiz, et al., 2012) and this study examined the potential for using participant response times as an alternative (Chen et al., 2016) by comparing data from both approaches.

### **4.2.2 Spatial and Visualisation Abilities**

As discussed in the literature section, multimedia research often includes participant measures of spatial ability as there is some evidence linking comprehension and retention performance with the wider cognitive spatial abilities of the reader (Gyselinck, Meneghetti, De Beni, & Pazzaglia, 2009; Meneghetti, De Beni, Pazzaglia, & Gyselinck, 2011; Meneghetti, Fiore, et al., 2011; Meneghetti, Gyselinck, Pazzaglia, & De Beni, 2009; Meneghetti, Labate, Grassano, Ronconi, & Pazzaglia, 2014; Plass, Chun, Mayer, & Leutner, 2003). In respect of graphical representations, evidence of such links has been found in studies involving weather maps (Allen, Miller Cowan, & Power, 2006), general maps (Pazzaglia & Moe, 2013; Sanchez & Branaghan, 2009), geospatial thinking (Huynh & Sharpe, 2013), science (Sanchez & Wiley, 2014), and programming (Anvari, Tran, & Kavakli, 2013), amongst other subjects. However, when considering a scientific diagram with a distinct direction of flow (an evolutionary sequence of dinosaurs), something which is similar to the flow characteristics of most timelines, Winn (1982) found that the participant's spatial ability did not directly predict learning, merely the ability to learn in the condition where the flow was contrary to normal reading directions.

The term 'spatial ability' is very broad, Colom, Contreras, Botella, and Santacreu (2002, p. 904) note that it is generally considered to be defined by the "generation, retention, retrieval, and transformation of visual images" and, from reviews of studies, that it has been decomposed by researchers into a wide variety of factors. Carroll (1993), as part of a comprehensive review of factor analytic studies of human cognition, identified common views regarding five aspects of spatial ability: visualisation, rotation, closure (mentally completing a shape from a partial image), flexibility of closure (identifying patterns in noise), and perceptual speed. D'Oliveira (2004) quotes a range of two to ten different factors across studies, while Hoffler (2010), in a meta analytical review of test instruments, points out that there are also a range of different labels applied by researchers to each factor. Linn and Petersen (1985) also found no consensus in categorisation of measures of spatial ability.

This study incorporated two measures of spatial ability: vividness of imagery (Marks, 1973; Andrade, May, Deeprose, Baugh, & Ganis, 2013), and mental rotation (Shepard & Metzler, 1971), which are separable (Dean & Morris, 2003), but related (Logie, Pernet, Buonocore, & Sala, 2011). In the Shepard and Metzler mental rotation task, participants are shown pairs of perspective drawings of three dimensional regular shapes made of blocks. The participant decides in each trial, whether the two shapes are identical but possibly rotated, or if one is a mirror-image of the other and could never be rotated and superimposed on the other.

Findings in the original research (Shepard & Metzler, 1971) and in subsequent studies (Cooper & Shepard, 1973; Brosnan, Walker, & Collomosse, 2010; Paschke, Jordan, Wustenberg, Baudewig, & Leo Muller, 2012; Shepard & Metzler, 1988; Provost & Heathcote, 2015; Tagaris et al., 1997) show response times linearly proportional to the angle of rotation between identical shapes, in line with the cognitive processing involved being one of mental rotation and that the slope and intercept of a regression line for the measured response times and angles can be used to characterise participant spatial ability in this aspect.

When considering whether recall of a timeline will assist the recall of associated information, whether through the CTML (Mayer, 2014a) or through conjoint retention (Kulhavy et al., 1985), the ability to visualise the timeline when it is no longer present may prove an important factor. A commonly used measure of the ability to recall images is the Vividness of Visual Imagery Questionnaire (VVIQ) (Marks, 1973) and the later version (VVIQ-2) (Marks, 1995). This study uses the visual element of a similar, but shorter, vividness questionnaire, the Plymouth Sensory Imagery Questionnaire (Psi-Q) which has been shown to be consistent with the VVIQ-2 (Andrade et al., 2013). The visual element of this questionnaire asks the participant to rate how strongly they can visualise a number of common images.

### 4.2.3 Aims

This study sought to identify any significant comprehension and learning differences between chronological information presented in text form and presented graphically in the form of timelines. It examined both initial comprehension/understanding and level of immediate retention using measurement of accuracy in tests and also assessed response times to allow a methodological comparison to be made between use of response time measurements as an indicator of cognitive load and the self reported instrument used by DeLeeuw and Mayer (2008).

The study addressed four (of the five) hypotheses numbered as in section 2.5 on page 55:-

Hypothesis 1: Use of timelines will result in increased comprehension of information when compared to text.

Hypothesis 2: Use of timelines will result in shorter question response times during comprehension of information when compared to text.

Hypothesis 3: Use of timelines will result in greater retention of information when compared to text.

Hypothesis 5: Question response times will correlate positively with self reported estimates of cognitive load.

As in similar research, discussed in the literature review, studies into multimedia have often included measures of spatial ability and perceived vividness of visual imagery, this study also included these in an effort to establish if there is a link between individual performance on these and performance in the two different presentation methods (text, timeline). In addition the study examined aspects of individual differences, both demographic (age, gender, education, timeline familiarity) and graphical/spatial ( visual verbal preference, and multimedia learning preference) to identify any correlation with the comprehension and retention results.

## 4.3 Method

### 4.3.1 Participants and Recruitment

The study participants were recruited through opportunity sampling of attendees during four open days at a south west England university held during 2015. At these events prospective students and their parent/guardians were offered the opportunity to experience a wide variety of experiments, including this study. A number of participants (63) had to discontinue the study part way through in order to attend other open day events. Across the 4 days however, a total of 131 finished the study. However, although all of those participants completed the text vs timeline story comprehension and retention questions, 10 of the participants chose not to answer one or more of the effort rating and preference questions (all questions were optional). Two groupings were therefore used for analysis. The main group, containing all participants, was used for the text vs timeline comparison, while the smaller subset, excluding those with incomplete effort/preference responses, was used for the remainder. The demographic, and highest education level summaries for the main set and the smaller subset are shown in Tables 4.1 and 4.2 and the relevant group is identified in each of the subsequent analyses.

Group	Number	Age Range	Mean Age	SD Age	Female
Main	131	16 - 58	27	15	84 (64%)
Subset	121	16 - 58	28	15	77 (64%)

**Table 4.1** – Study Two: Demographics

Group	No qualifications	GCSE	A level	Diploma	First Degree	Postgraduate
Main	2	29	68	7	13	12
Subset	1	28	61	7	12	12

**Table 4.2** – Study Two: Participant Highest Education Level

## 4.3.2 Materials and Measures

### 4.3.2.1 Experiment framework

The experiment was performed via a computer based questionnaire using the open source Limesurvey software (Limesurvey, 2015, Version 2.05+ Build 141210) , which was modified to provide response time measurements for all relevant question types (the standard software only covers simple response types such as "yes/no", and "list selection"). Although it is possible to produce the questionnaire pages one at a time, it was clear that, to cope with the number of pages involved (308 distinct pages), and also ensure consistency, an automated, scripted approach was needed. The pages were therefore produced by a set of scripts, using standard texts/images, before being uploaded into the researcher's Limesurvey server. The random number function of the Limesurvey software was used to counterbalance groups in the comprehension phase by selecting the appropriate route through the questions. Initial trials of the spatial rotation tests showed that the response time measurement capability built into Limesurvey suffered from a large variation in measurement accuracy. A controlled trial using an automated test rig with a fixed 1000ms stimulus response resulted in widely varying measurements ( $M = 1938ms$ ,  $SD = 301ms$ ). Clearly, when it is considered that Shepard and Metzler (1971) found mental rotation speeds of around 60 degrees per second, while Ganis and Kievit (2015) quote 100 degrees per second (albeit for a slightly different set), a measurement error distribution with an  $SD$  of 301  $ms$  would have a significant impact upon the reliability of measurements that are themselves around 400ms. A javascript program suitable for embedding into each page of the rotation timing tests was therefore developed (and incorporated through the scripting process described earlier). The same automated trial of this approach resulted in an improved measurement accuracy and reliability ( $M = 1069ms$ ,  $SD = 30ms$ ).

#### **4.3.2.2 Overall Experiment structure**

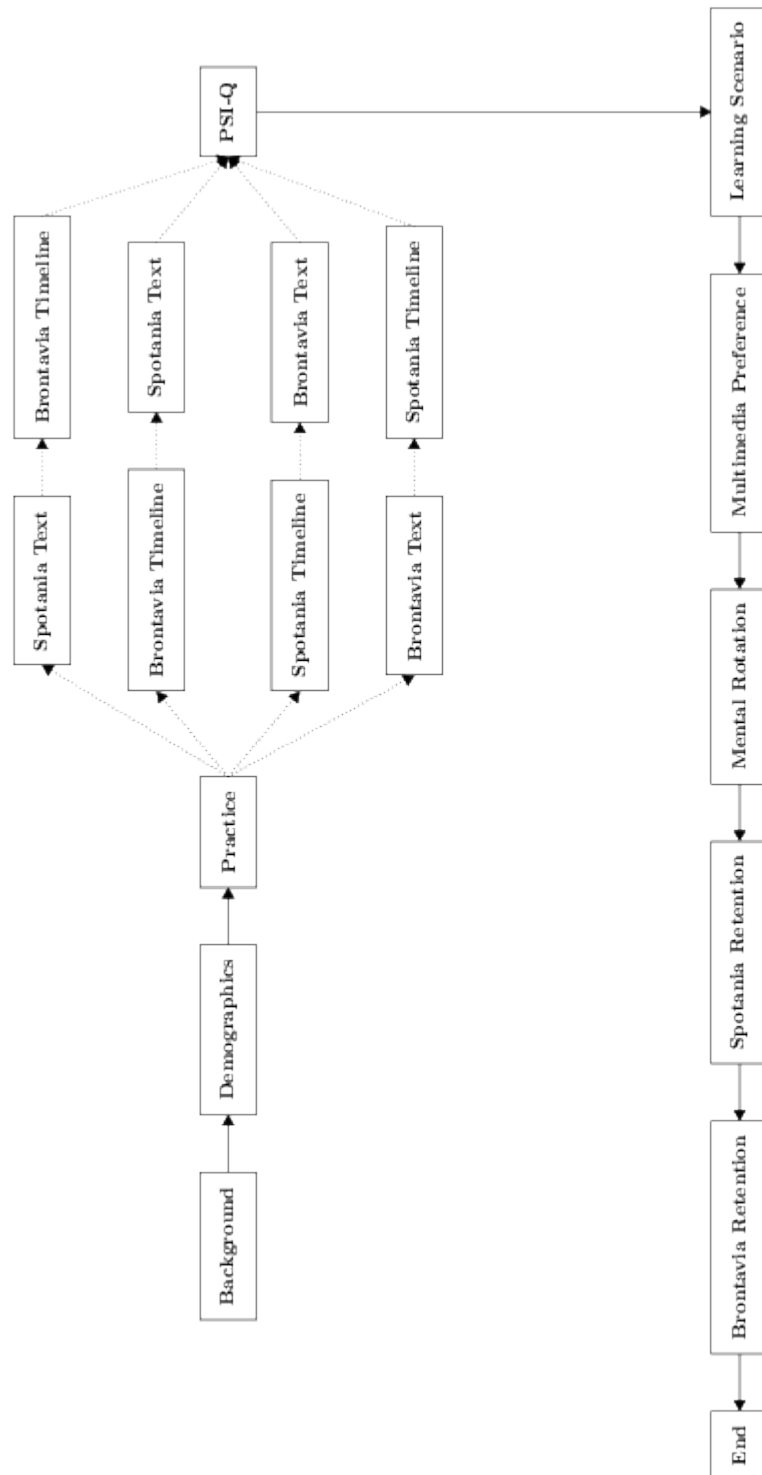
To meet the aims of the study a wide variety of materials were used and linked together in the sequence shown in Figure 4.1. In the first (comprehension) phase the sequence of materials and tests was: background information, demographic questions, practice examples of each of the three question types, study of the first story, questions on the first story, study of the second story, questions on the second story.

#### **4.3.2.3 Text and timeline materials**

Consistent with the findings of study one, the materials used for the present study were based around history as the subject matter. In order to prevent prior knowledge being a confound, completely fictional materials were written by the researcher, in a similar manner to that used by Voss and Silfies (1996), and a light, slightly humorous, style was chosen in order to maintain participant interest (Zabidin, 2015; Ferstl, Israel, & Putzar, 2017).

Generally similar in structure, the stories can be seen in Appendix C on page 241. The texts were balanced as closely as possible in terms of number of dates, names, and causal linkages etc. and in reading comprehension scores Flesch-Kincaid (Kincaid, Fishburne Jr, Rogers, & Chissom, 1975) and Flesch Reading Ease (Flesch, 1948) (both as reported by the word processing tool, Microsoft Word, Professional Office Plus). Both texts were rated as difficult (Farr, Jenkins, & Paterson, 1951) but it should also be noted that automated tools take a simplistic view of text complexity (Schrivver, 1989, 2000). A summary of the analysis of each story can be seen in Table C.1 on page 244 showing for each story the number of names and dates included (in both text and timelines), the number of distinct lines in each timeline, the number of words and the reading comprehension scores for the text versions. The stories contained sufficient information for the various questions.

The key points from the text were incorporated into XML representations and graphical timelines produced from those, using the Simile tool from MIT as described by Butler,



1

**Figure 4.1** – Study Two - Experiment structure



Gilbert, Seaborne, and Smathers (2004), with the resulting images captured, by a screen capture tool, for incorporation into the questions. The timelines can be seen in Figures C.1 on page 245, and C.2 on page 246. The event labels were placed in the appropriate locations minimising split attention (Florax & Ploetzner, 2010; Kalyuga et al., 1999; Sweller et al., 2006), and following the associated spatial contiguity principle (Mammarella et al., 2013; Huff, Bauhoff, & Schwan, 2012).

#### **4.3.2.4 Comprehension and immediate retention questions**

The questions were of three types: numerical (two digit/four digit), multiple choice selection (one of two, three, or four choices), and the placing in order of three items. Examples of the question types used in the comprehension phase can be seen in Figures C.3, C.4, C.5, C.6, and C.7 on pages 247 to 255. The retention phase used the same layout, but without the material being shown, as can be seen in the example in Figure C.14 on page 260. Table C.2 on page 248 lists the content and style of each of the comprehension questions, while Table C.4 on page 250 shows those for the retention phase.

#### **4.3.2.5 Self reported effort and difficulty**

Self-rating of mental effort via a 7 point Likert scale (1 = extremely low, 7 = extremely high) was included after one example of each question type in each group for the comprehension phase (arranged so that the participant was never asked to rate this on consecutive questions) together with an assessment of overall difficulty of each group in the comprehension phase using a 5 point Likert scale (1 = extremely easy, 5 = extremely hard) included after completion of each story. Both scales are in line with the approach used by DeLeeuw and Mayer (2008) (to indicate intrinsic and germane cognitive loads respectively).

#### **4.3.2.6 Plymouth Sensory Imagery Questionnaire (PSI-Q)**

The visual related portion of the PSI-Q as described by Andrade et al. (2013) was incorporated, and can be seen in Figure C.8 on page 256.

#### **4.3.2.7 Learning Scenario Questionnaire (LSQ)**

The instrument developed by Mayer and Massa (2003) was incorporated in a set of questions in the style shown in Figure C.9 on page 257. The complete set of questions is shown in Table C.5 on page 251.

#### **4.3.2.8 Multimedia Learning Preference Questionnaire (MLPQ)**

Mayer and Massa (2003) also developed an instrument to assess multimedia learning preferences which was included as a set of three questions that can be seen in Figures C.10, C.11, and C.12 on pages 257 to 259.

#### **4.3.2.9 Mental rotation ability**

The materials for the Shepard and Metzler mental rotation task (Shepard & Metzler, 1971), where participants are shown pairs of perspective drawings of three dimensional regular shapes made of blocks, were obtained from a set published by Tarr (2015). The pairs were incorporated into displays in the format shown in Figure C.13 on page 259. There were 16 randomised pairs, eight were congruent and eight incongruent, with the congruent pairs including each rotation angle of 40, 60, 80, 100, 120, 140, 160, 180 degrees.

#### **4.3.2.10 Analytical approach**

A program was developed to score the outputs from the various parts of the study; the methods used are described below:

**Main questions (comprehension and retention)** The multiple choice questions were scored with a value of one for the correct answer. Numeric questions were scored the same way, a mark being awarded only if the answer was exactly correct, as this is typical of school examinations, which have a prescribed mark scheme. To ensure data were not overlooked, the results were checked to identify any particularly popular wrong answers. Similarly, although it could have been possible to still score the ordering questions if only two out of the three/four items were correctly ordered, the simpler approach of only allocating a value of one for a full correct ordering was chosen. Informal discussion with the teachers during study one had suggested that this would commonly be used as an assessment approach in marking at a school. Had the ordering questions contained a larger number of items however, then an approach that scored sub groups of correctly ordered but displaced items as used by Foreman et al. (2008) would have been used, but the increase in scoring complexity for the very few cases (two cases out of six) was not considered worthwhile. The scores were converted to percentages of the maximum score for each phase (12 for comprehension, six for retention). The scores were prepared for analysis by grouping into text or timeline conditions of the study materials.

**Vividness** The mean of all the responses to the questions in the PSI-Q subset was used as the outcome, in line with Andrade et al. (2013).

**Learning scenario preference** In line with Mayer and Massa (2003) a score of one was allocated for answers where the visual mode was preferred and the results were totalled to give an overall result in the range zero to five.

**Multimedia preferences** Also in line with Mayer and Massa (2003) a score of one was allocated for each answer where the preference was for the visual help screen and the results were then totalled to give an overall result in the range zero to three.

**Mental rotation** The measure used from this test consisted of the angles and response times for correct responses which were processed in line with Shepard and Metzler (1971), and Brosnan et al. (2010), i.e. discarding incorrect responses and using the slope and intercept of a regression line fitted to the data as the resultant measures.

### 4.3.3 Design

The study involved multiple elements and had three distinct phases: First an examination of comprehension accuracy and response times for textual and timeline material, secondly a similar assessment of accuracy and response times concerning participant retention of the same material, and then a third phase collecting demographic information, and a range of data on spatial/visualisation abilities and preferences. The latter being interposed between the assessments of comprehension and retention to act as distractor tasks preventing any rehearsal in working memory of the stories, before the retention questions (Hinze, Wiley, & Pellegrino, 2013). The comprehension phase was a mixed within-between design, with the independent variable being timeline vs text presentation and two dependent variables (accuracy and response time). The design was counterbalanced with the participant being randomly allocated (by the software) to one of four sets of questions (text first x timeline first, starting with Story 1 x starting with Story 2). During the comprehension phase the material remained visible while the questions were being asked. The retention phase (for which the material was not shown) involved a set of questions regarding each story, in a fixed order. The ‘within participant’ variables for the comprehension and retention portions were participant accuracy and response times for each question. The ‘between participants’ variables were age, gender, highest level of education, occupation, usage of timelines, spatial rotation ability (Shepard & Metzler, 1971), self rated vividness of visual imagery (Marks, 1973; Andrade et al., 2013), and Multimedia and Multimedia Learning Preferences (Massa & Mayer, 2006). The placement of each item can be seen in Figure 4.1 where the counterbalanced routes through the comprehension phase can be seen. The mental

rotation and multimedia preference assessments, etc. were placed between the comprehension and retention phases to also act as distractor tasks to prevent rehearsal of the fictional stories within working memory before the retention questions.

#### **4.3.4 Procedure**

The study took place in a computer suite adjacent to the main laboratory where the open day was taking place so that the environment was relatively quiet. After a welcome by the researcher and allocation to a computer, participants were given time to study the background information, obtain answers to any questions they had, and to sign a consent form. The questionnaire was viewed via a standard browser (Google Chrome) with the participant able to adjust the window size up to the maximum of the screen resolution of 1366x768 on a 20 inch Liquid Crystal Monitor. The first web pages of the study provided background information before the participant was asked by the software to enter their demographic information. The next pages provided an explanation of the experiment and incorporated practice examples of three of the basic question types (numeric, selection, and ordering). The main questions followed, with two distinct groups, one of which used a timeline, and the other text. The first page in each group provided the story in text or timeline form and the participant was allowed to take as long as they wished to study it (within the overall period of the open day). They were also told that the information would be repeated on every subsequent page of the comprehension test. For the subsequent pages containing questions, the participant was asked to answer as quickly and accurately as they could. There was a clear division between the sets of questions for the two different stories, with a similar opportunity to study the text/timeline, on its own, provided before the second set of questions started. The next set of questions used self report to assess Vividness of Visual Imagery (Marks, 1973), Learning Scenario, and Multimedia Learning Preference (Massa & Mayer, 2006; Mayer & Massa, 2003). The participant was then presented with pairs of shapes, as used by Shepard and Metzler (1971), and asked to identify

whether these were the same or different. After completing the retention questions the participant was thanked for taking part, and provided with a debrief if requested.

## **4.4 Results**

### **4.4.1 Group balance**

The groups did not differ significantly in terms of their proportion of gender,  $\chi^2(3) = 2.48, p = .48$ , nor in terms of age,  $F(3, 117) = .498, p = .684, \omega = -.01$

### **4.4.2 Participant Accuracy**

Examination of answers given to the numerical questions showed no common errors (something that might otherwise have cast doubt on the approach of only awarding the mark for a correct answer). The four distributions of total marks for the two conditions (text vs timeline material) and the two measurement points (initial comprehension and subsequent retention) were checked for normality using Shapiro-Wilk tests since this test was suggested by both Thode (2002) and Razali and Wah (2011). The detailed results can be seen in Table H.1 on page 276. All four distributions were found to be significantly different from a normal distribution ( $W$ 's  $< 0.92$ ,  $p$ 's  $< .05$ ). The distributions were also found to be skewed, with inconsistent direction preventing use of a transform since the same transform would need to be applied to each in order to be able to perform valid comparisons (Field et al., 2012). With such data Larson-Hall and Herrington (2010), and Erceg-Hurn and Miroseovich (2008), suggest the use of robust methods, and Wilcox (1992) stresses concerns with power and relevance if they are not used. Comparison of the marks for the comprehension test showed that the distribution of marks for the text group with a 20% trimmed mean ( $M = 80.25$ ), did not differ significantly from that for the timeline group, with a 20% trimmed mean ( $M = 82.88$ ) when assessed using Yuen's

method,  $T_y(72) = -1.76$ ,  $p = .083$ , 95%  $CI[-5.6, 0.35]$  with an explanatory measure of effect size ( $\hat{\xi} = 0.12$ ), roughly corresponding to a Cohen's d value of 0.2 (Wilcox, 2016; Ellis, 2010). However comparison of the marks for the immediate retention test showed that the distribution of marks for the text group with a 20% trimmed mean ( $M = 63.24$ ), did differ significantly from that for the timeline group, with a 20% trimmed mean ( $M = 55.71$ ) when assessed using Yuen's method,  $T_y(72) = 2.76$ ,  $p = .007$ , 95%  $CI[2.1, 12.97]$  with an explanatory measure of effect size ( $\hat{\xi} = 0.26$ ), roughly corresponding to a Cohen's d value of 0.35 (Wilcox, 2016; Ellis, 2010).

#### **4.4.3 Participant Response Times**

The response times for the questions ranged from 1.5s to 168s (one outlier at > 700s was replaced with the next highest value of 168s) with  $M = 25.8$ ,  $SD = 20.3$  and were, as predicted by Whelan (2010), non-normally distributed, with skewness for the whole set of measurements of 2.0 ( $SE = 0.38$ ) and kurtosis of 5.6 ( $SE = 0.38$ ). However, since some of the types of questions involved different actions by the participant (for example age and date questions required use of the keyboard, and order questions involved greater use of the mouse than simple selection) and, since the cognitive load may also differ between types, the overall comparisons of response times were only considered appropriate when made between similar questions. Examining the response times for each question also showed evidence of a speed vs accuracy trade off (Anderson, 1981). The response time comparisons between timeline and text conditions were therefore restricted to the cases where the participant had given correct answers in both conditions.

**Comprehension Question Response Times** The distributions were compared using Yuen's method for paired distributions (Wilcox, 2016). Significance values were adjusted using Holm's method (Holm, 1979) and comparisons for each question can be seen in Table H.5 on page 277. For five of the comparisons the differences are significant, and in

each case the timeline case has the lower trimmed mean. These cases are reported here: For question two, the 20% trimmed mean time of 27.27s for the timeline condition was faster than the text condition which had a mean of 35.86s

$T_y(54) = 3.37, p = .014, 95\% CI[3.48, 13.71]$ . For question three, the 20% trimmed mean time of 13.04s for the timeline condition was faster than the text condition which had a mean of 19.03s  $T_y(55) = 4.79, p < .001, 95\% CI[3.48, 8.5]$ . Question seven resulted in 20% trimmed mean time of 14.63s for the timeline condition which was faster than the text condition with a mean of 17.38s,  $T_y(63) = 2.86, p = .045, 95\% CI[0.83, 4.68]$ , while question eight resulted in 20% trimmed mean time of 12.46s for the timeline condition which was faster than the text condition with a mean of 20.22s  $T_y(50) = 4.88, p < .001, 95\% CI[4.57, 10.95]$ . Finally, question 12 resulted in a 20% trimmed mean time of 8.54s for the timeline condition which was faster than the text condition with a mean of 9.98s  $T_y(69) = 3.3, p = .014, 95\% CI[0.57, 2.3]$ .

**Retention Question Response Times** Using the same approach of comparing response times for questions where both answers were correct for the immediate retention test resulted in the comparisons shown in Table H.6 on page 278. It should be noted, however, that relatively small numbers resulted for questions two to five. Question one however did show a significant difference with, in this case, the 20% trimmed mean time of 7.25s for the timeline condition being slower than the text condition with a mean of 6.31s

$T_y(64) = -2.82, p = .038, 95\% CI[-1.61, -0.28]$ .

#### 4.4.4 Other Variables

The participant preferences in respect of Multimedia Preference (MMP), Multimedia Learning Preference (MLPQ) (Massa & Mayer, 2006) and vividness of visual imagery (Andrade et al., 2013; Marks, 1973) were collected. These, together with the demographic variables of gender and age, slopes and intercepts for the mental rotation task,



	<i>Age</i>	<i>MRS</i>	<i>MRI</i>	<i>MMP</i>	<i>LSP</i>	<i>VVT</i>	<i>CompD</i>
MRS	-0.13						
MRI	0.29**	-0.37****					
MMP	0.12	0.00	0.01				
LSP	0.15	0.10	-0.13	0.21*			
VVT	0.03	0.19*	0.08	-0.02	-0.01		
CompD	-0.12	0.03	-0.07	-0.03	0.07	-0.12	
RetnD	-0.27**	-0.05	-0.11	-0.08	0.00	-0.11	0.19*

MRS = Metzler Slope, MRI = Metzler Intercept, FRS = F Rotate Slope, FRI = F Rotate Intercept, MMP = Multimedia preference, LSP = Learning Scenario Preference, VVT = Vividness from PSI-Q, CompD = Comprehension score diff, RetnD = Retention score diff, \*P < .05, \*\*P < .01, \*\*\*P < .001, \*\*\*\*P < .0001,

**Table 4.3** – Study Two Correlation Table

and the differences between participant scores for the text and timeline conditions in comprehension and retention were examined for any correlations by means of Pearson tests. The resulting matrix can be seen in Table 4.3. The only significant correlations were between broadly equivalent measures such as the two mental rotation tests and between the comprehension and retention scores (repeated measures of the same participants). There were therefore no meaningful strong correlations found. The variables were also used to group the accuracy scores and response times for analysis with no consistent, significant, outcomes.

#### **4.4.5 Comparison of the two approaches to cognitive load measurement**

Two measures were used to assess cognitive load, the response times to each question, and self assessed effort questions as used by DeLeeuw and Mayer (2008). To compare the two approaches, correlations were performed, using the Spearman method, between the participant assessed effort rating and the response time recorded for the question to which

	All Text		All Graphics		Correct Text		Correct Graphics	
	Rho	P	Rho	P	Rho	P	Rho	P
Number (2 digits)	0.20	0.03	0.29	0.00	0.22	0.04	0.38	0.00
Order	0.41	0.00	0.42	0.00	0.55	0.00	0.37	0.01
Choice (1 of 4)	0.50	0.00	0.46	0.00	0.46	0.00	0.51	0.00
Number (4 digits)	0.29	0.00	0.19	0.03	0.33	0.00	0.19	0.05
Choice (1 of 2)	0.30	0.00	0.46	0.00	0.30	0.00	0.47	0.00

**Table 4.4** – Study Two: Effort and Response Time Correlations for Different Question Types

that effort rating applied. The values of Rho and p for the complete set of results for text and data and also for the subset where the question was also correctly answered can be seen in Table 4.4. It can be seen that whilst some correlations that achieve significance are weak, there are some that can be considered as moderate and all significant correlations are in the same direction. The correlations in respect of the one of four choice question are consistently the highest.

**Difficulty ratings** At the end of each of the two sets of questions (timeline condition and text condition) a self rated difficulty rating was provided. The responses were highly skewed (towards the easy end of the scale) and so comparison of these responses between conditions was performed using Yuen’s method for paired distributions. No significant difference was found between the timeline condition with a 20% trimmed mean ( $M = 1.37$ ) and the text condition with a 20% trimmed mean ( $M = 1.44$ ),  $T_y(72) = 0.57, p = .571, 95\% CI[-0.17, 0.31]$ .

#### 4.4.6 Overall Timing

Overall timings were recorded to assist the design of some of the subsequent studies by providing indications of the range of reading speeds for the material. The fastest participant

took 13 min to complete the whole study, the slowest 67 min ( $M = 30 \text{ min}$ ,  $SD = 11 \text{ min}$ ). The familiarisation element of the comprehension test was studied for a mean of 88s ( $SD = 64 \text{ s}$ ) with less time taken on average for familiarisation with the second set of material ( $M = 43 \text{ s}$ ,  $SD = 47 \text{ s}$ ). Text stories were studied for a mean of 86s ( $SD = 68 \text{ s}$ ), and timeline versions for a mean of 44s ( $SD = 42 \text{ s}$ ).

## 4.5 Discussion

As discussed in Chapter two there has been little work into the benefits of timelines for supporting learning, with a major gap being whether timelines can empirically show a learning benefit. When considering learning and timelines it was postulated that they could be considered as multimedia, combining graphics and text, and hence that they should help with learning if they were constructed following the principles described in the CTML. While study one provided a first step towards understanding timelines and education, and indeed in the year 8 group suggested that the use of timelines did improve the learning (as assessed by a comprehension based test), a number of limitations were found, mostly resulting from being conducted in a school and prioritising ecological soundness. Therefore a key aim of this second study was to examine timelines and learning in a more controlled scenario. Of the five hypotheses being examined by the set of studies within this thesis, four of these were addressed by this study. Two concerning the effects of timelines on comprehension (in terms of accuracy and response time), one extending the assessment of accuracy to cover immediate retention, and the fourth using the testing process (independent of the type of material involved) to ascertain the extent to which response times could provide a measure of cognitive load.

Designed to take place in a controlled lab environment, this study benefited from the introduction of fictional materials to prevent prior learning effects (Ozuru et al., 2009; Lowe, 1996), using two broadly identical, history themed, stories, one in text form and the other as

a timeline.

Previous timeline studies by Korallo (2010b) and Prangma (2007) had both focused on retention, showing initial improvements from inclusion of timeline information, although both showed no difference for delayed retention. To build on this work in a stepwise manner, this study was designed to focus on the immediate retention and the underlying comprehension needed to support this.

Consideration of the CTML (Mayer, 2005), and its foundations in DCT (Paivio, 1991), and CLT (Sweller, 1988, 1994) predicted that, as examples of multimedia, the use of timelines would be beneficial for retention of the information that they were presenting, and, since comprehension is a necessary precursor to retention (Schnitz & Bannert, 2003; Kintsch, 1994), it was expected that the same effect would be found when comparing comprehension. The results, however, showed no significant difference in accuracy between conditions (text vs timeline) at the comprehension stage, but found a significant accuracy difference when assessing immediate retention, with the trimmed mean accuracy score for the text condition being higher than that for the timeline. These were both unexpected results. In particular the retention result which contradicted previous work by Korallo (2010b) and Prangma (2007). A second facet of this study was to explore the measurement of cognitive load. While considering learning and its underlying theories, the literature section, chapter two, discussed theories of memory (Baddeley & Hitch, 2000; Atkinson & Shiffrin, 1968; Cowan, 2014) and the effects of the cognitive load experienced during learning (Sweller, 1988). In particular Sweller's studies of Cognitive Load Theory (Sweller, 1994, 2015) suggested that cognitive load induced by study materials can be a good indicator of how well they will be learned. Research has used a number of different methods for measuring cognitive load, but the commonly used method of self report interrupts the participants as they take part in the learning process. Consequently this research was also interested in the extent to which reaction times could give an indicator of cognitive load (Chen et al., 2016) and so also included a comparison of reaction time and self report cognitive load assessment (van Gog

& Paas, 2008). Correlations between measured and self-reported cognitive load were all positive, and significant in all but one example, but showed either weak or moderate strengths. The measurements were grouped by type of question (since underlying response times differed between those involving typing and those involving selection by mouse click for example) and the strongest correlations were seen for those involving simple, mouse based, choices, where the underlying timing variance would have less of an obscuring effect. This study also sought to ascertain whether factors such as education, age (Borella et al., 2011), spatial ability (Sanchez & Wiley, 2014; Pazzaglia & Moe, 2013), and gender (Coluccia & Louse, 2004) could affect the measurements being made. Individual differences including gender, age, highest education level, occupation, and familiarity with timelines, showed no significant correlations with either accuracy or cognitive load, nor did assessments of spatial ability, visualisation, and learning preferences, whether directly or by using these as grouping selectors for the main accuracy and response time variables. In the case of spatial ability there are many possible aspects and different tests of these, Sanchez and Wiley (2014) for example, only found effects using a dynamic spatial ability assessment and not with a paper folding approach (which is perhaps closer to the mental rotation assessment used here).

In summary neither hypothesis one nor three were supported, with timeline materials showing no evidence of the predicted difference in comprehension, or immediate retention, and interestingly indicating the opposite effect to hold in respect of hypothesis three concerning immediate retention. When reaction times were used as a measure of ease of comprehension however, in line with hypothesis two, the results did support the hypothesis in the cases where the results were comparable. Correct responses from participants were significantly faster for questions regarding timeline material than those for the text format. Finally, hypothesis five was supported and, it is concluded that it may be possible, in similar studies, to use response time as a measure of cognitive load instead of the more intrusive self assessment approach. It should be noted, however, that to be meaningfully comparable, the

analysis process necessitated a reduction in the directly compared data at a number of stages (only comparing similar question types since their distributions varied, and only comparing correct answers, since correct/incorrect distributions also differed). As a consequence, studies using the approach to gain insight into the load for individual participants would need to include a larger number of questions than this study did which was only examining correlation between the two methods of estimating cognitive load and hence was able to aggregate the measurements.

As the results of this study were mixed, however, this next section looks to understand and reflect upon factors, which may have impacted the study. Considering the different results for hypothesis one between study one and two, the primary measurement approach used in study one was based upon the educational grading system used by the school, and incorporated other data (previous scores and targets) into the analysis. Hence it could have been more sensitive than the single test score used for comprehension in study two.

To develop this study further and provide insights into the participants, the next study was designed to include more tests of mental rotation, with the participants using the keyboard rather than the mouse to provide greater timing consistency since, on examining the wide variation in times for the mental rotation tests, it had been realised that some of the timing variation would have been caused by moving the mouse to the correct answer and more accurate results would be needed to be certain that there was a difference. The incorporation of a further delayed retention phase was intended to help assess any longer term effect on learning (in case the relative immediacy of the retention test in the current study made a difference).

## **Chapter 5**

# **Study Three: An experimental study on the use of timelines with delayed assessment**

### **5.1 Chapter Overview**

Building on study two, this study also aimed to identify comprehension and retention differences between chronological information presented as text and presented graphically in the form of timelines. While keeping the same initial materials, structure, and approach as study two, the number of measurements of spatial rotation ability was extended to improve accuracy, and an additional spatial ability test of object location memory was incorporated. The assessment of retention was enhanced by the inclusion of a delayed retention test a week later. As had been found in study two, there were no significant differences in the distribution of accuracy results for comprehension between timeline and text materials. Unlike study two, however, there was also no significant difference found for the results relating to the initial retention questions. A significant difference was found for accuracy in the delayed retention session reflecting a higher trimmed mean score for the material that

had been originally studied in text form. Examination of response time distributions during the immediate retention test comparing identical questions for the two material types resulted in only one significant difference. This was, however, in the opposite direction to those found in comprehension (and in study two), with the trimmed mean for the text condition showing the faster response.

## **5.2 Introduction**

### **5.2.1 Rationale**

The key finding from study two was that while there were no significant differences in accuracy scores for assessments of participant comprehension between material presented as a timeline and as text, there was a small, but significant, difference favouring textual presentation, during a retention test shortly after the comprehension study of the materials.

This was unexpected and contrary to predictions based upon the literature.

Study two also sought to understand cognitive load and whether response times (RT) could be used as an indicator of cognitive load in comparison to the typically used self reporting as assessed through an instrument developed by Massa and Mayer (2006). The results indicated significant correlation coefficients (varying between weak and moderate for different questions) indicating that RT could be used as an alternative to self report.

Comparing the response time results for timelines and text conditions no significant results were found for the retention session but in those questions where significant differences were seen for the comprehension questions, the timeline response times were shorter than the corresponding text condition.

Study two was also designed to be run in a single experimental session and, whilst intervening tests (mainly of spatial ability) provided a break to remove any working memory rehearsal effects between the comprehension and retention elements, results that would be more meaningful for learning and education would require a delayed retention test as this is



more representative of the experience learners would have in schools with testing typically occurring in a later lesson, often the following week. Therefore in this next study an additional retention test was introduced, designed to occur one week ( $7 \pm 2$  days) after the initial exposure and testing of the material.

An additional advantage of including a delayed test in this third study was that it also allowed for the inclusion of additional spatial ability tests. By testing over the two sessions it not only meant that more data could be collected in relation to spatial awareness, but it also ensured that the score was more representative of the participant, having been collected over two sessions (a participant having an "off day" may lead to unexpected results). A test of object location memory was also incorporated. These additions to enhance and extend the spatial ability aspect of the study, and the delayed retention test, are described in more detail below. All other aspects are the same as for study two.

#### **5.2.1.1 Delayed retention test**

Amongst memory and education researchers reference is frequently made to research in 1885 by Hermann Ebbinghaus as described in Ebbinghaus (1885). As part of his research he studied lists of nonsense syllables until he could recall them accurately on two consecutive tests and then measured his retention by tests performed over a range of intervals from 20 minutes to 31 days (Hicks, Marsh, & Russell, 2000). Many subsequent studies have been performed with a variety of memory aspects being assessed, from sequences, nouns (van den Broek, Segers, Takashima, & Verhoeven, 2014; Wheeler, Ewers, & Buonanno, 2003), word pairs (Soderstrom & Bjork, 2014), prospective actions (Hicks et al., 2000), to pictures (MacLeod, 1988; Hockley, 2008). The studies have also used a wide variety of durations from minutes to years. For example, Chan (2010) uses 20 minutes/one day/one week, Kang, McDermott, and Roediger (2007) consider two days, three days and one week, Roediger and Karpicke (2006) use five minutes, two days, and one week. Rubin and Wenzel (1996) performed a meta level analysis of 210 data sets examining their fit to 105 possible

mathematical models for the rate of forgetting, concluding that the simplest (although with issues in boundary conditions) for common use was a logarithmic function. Many of these retention studies include accuracy feedback during the initial study session, engaging the testing effect (Roediger & Butler, 2011; Roediger & Karpicke, 2006; Carpenter, Pashler, & Cepeda, 2009) and the experiment materials were modified to provide similar feedback. The period of a week later than the first session was chosen for the second retention session as a reasonably practical period, with examination of the Ebbinghaus curve in Anderson and Schooler (1991, figure 1.A) showing a likely reduction of remembered items to 30% of the initial level.

#### **5.2.1.2 Broader assessment of spatial ability**

The Shepard and Metzler mental rotation test has been shown by some researchers to correlate with gender, with males in general performing better than females (Masters, 1998; Parsons, 2004; Gluck & Fabrizii, 2010; Doyle & Voyer, 2013) in both the intercept (showing faster fixed processing time), and slope, (indicating faster rotation processing time). In contrast, however, in a memory test developed by Silverman and Eals (1992) where participants memorise locations of objects and then identify subsequent changes of the objects, including absence and/or changes of their locations, a female advantage has been shown (Robert & Savoie, 2006; Silverman, Choi, & Peters, 2007; Lejbak, Vrbancic, & Crossley, 2009; Honda & Nihei, 2009). A version of this test was therefore incorporated, to complement the mental rotation tests.

#### **5.2.1.3 Increased assessment of mental rotation**

The addition of a second session also allowed for the Shepard and Metzler (1971) test used in study two to be extended. This test was intended as an exploratory indication of abilities that could be correlated with other variables and used to group other results. While the original task used 1600 image pairs, (Shepard & Metzler, 1971) , a review of the literature

indicated that researchers are not consistent in the number of pairs they use. Brosnan et al. (2010) obtained usable data from 40, Gardony, Taylor, and Brunye (2014) used 90, Amorim, Isableu, and Jarraya (2006), when comparing the effect of block shapes to human body shapes used 84 of each, Hugdahl, Thomsen, and Ersland (2006) used only 36 but were simply interested in accuracy and reaction times using the shapes as stimuli in conjunction with fMRI rather than evaluating rotation rates. Larger numbers, such as 128 were used by Borst (2013) (in two blocks with other conditions varied), 200 by Paschke et al. (2012), and 478 by Parsons (1987) (using abstract shapes). Clearly there is currently little consistency in how many pairs are required for effective use of the test. As this measure is used partially as a distraction and is not the key aim of the study, it was decided arbitrarily to use 120 pairs, split into two parts, with 60 in each of the two sessions.

The Shepard and Metzler mental rotation test was complemented by means of the incorporation of a second, different, rotation test, into each session. This used rotated letters rather than abstract shapes (Tarr & Pinker, 1989; Koriat & Norman, 1985) which some researchers had found to be less affected by gender (Beste, Heil, & Konrad, 2010). Pairs of the letter "F" were used (48 in each session). To provide additional balance, an additional spatial ability test, Object Location Memory was also introduced. In this test, developed by Silverman and Eals (1992), a female advantage has been shown (Robert & Savoie, 2006; Silverman et al., 2007; Lejbak et al., 2009; Cinan et al., 2007; Honda & Nihei, 2009).

### **5.2.2 Aims**

This research seeks to identify any significant comprehension and learning differences between chronological information presented in text form and presented graphically in the form of timelines. It seeks, in general, to measure both initial comprehension/understanding and the level of immediate retention. The study addressed three hypotheses (of the overall five) numbered as in Section 2.5 on page 55:

Hypothesis 1: Use of timelines will result in increased comprehension of information when

compared to text.

Hypothesis 2: Use of timelines will result in shorter question response times during comprehension of information when compared to text.

Hypothesis 3: Use of timelines will result in greater retention of information when compared to text.

## **5.3 Method**

### **5.3.1 Participants and Recruitment**

The study was granted ethical approval from the Department of Psychology ethics committee (ethics reference number:14-238). and participants were recruited through a variety of means: through the Psychology Department's Research participation Scheme (RPS), through general publicity on the university noticeboards, and via opportunistic sampling. All first year psychology undergraduates at the university are expected to participate, through the RPS, in at least five hours of research, split into half-hour blocks, and receive partial course credit for a Psychology Laboratory unit for completion of the five hours. No incentives were provided for any other participants apart from snacks being available during the experiment. All participants were assumed to be fluent in English, this being one of the entry requirements to study at the University. A total of 52 participants took part in the experiment, the age range being 18 to 43 years ( $M = 20.13$ ,  $SD = 4.39$ ) and the majority (44) falling into the 18-20 year range. The gender split was 41 female. Forty three of the participants had used timelines before and 40 made frequent use of maps/diagrams/graphs (at least once a month) in their job, studies or hobbies. The highest education level for the majority (42) was A-Level (with 2 holding diplomas, 6 having degrees, and 2 doctorates).

Group	<i>n</i>	Female	Age			
			Min	Max	Mean	SD
1	11	9	18	30	20	3.4
2	14	11	18	43	22	6.6
3	16	12	18	21	19	0.8
4	7	5	18	35	21	6.1

**Table 5.1** – Study Three - Group allocations

## 5.3.2 Materials and Measures

### 5.3.2.1 Experiment framework

The experiment was performed via a computer based questionnaire using the open source Limesurvey software (Version 2.05+ Build 141210) in the same manner as study two, the scripted approach being even more necessary to handle around 700 distinct pages across both sessions.

### 5.3.2.2 Overall Experiment structure

The experiment involved two separate sessions, the first being similar to that used in study two (with the only change being the addition of the letter rotation test and an increase in the number of Shepard and Metzler shape rotation pairs). The elements of the session can be seen in the overview of Figure 5.1. The second session, one week later ( $7 \pm 2$  days), commenced with questions on the text/timeline material that had been studied in the first session. Two blocks of questions were used, one for each story, but the choice of which was asked about first was decided at random (using the random number capability of the survey tool). The overview of the second session can be seen in Figure 5.2, where the placement of the two mental rotation tasks (shape rotation and letter rotation) and the object location memory task is shown. To preserve anonymity the only cross reference between the two sessions used a word or sequence of characters chosen by the participant and entered in both

sessions in the stages marked as "Link".

### **5.3.2.3 Text and timeline materials**

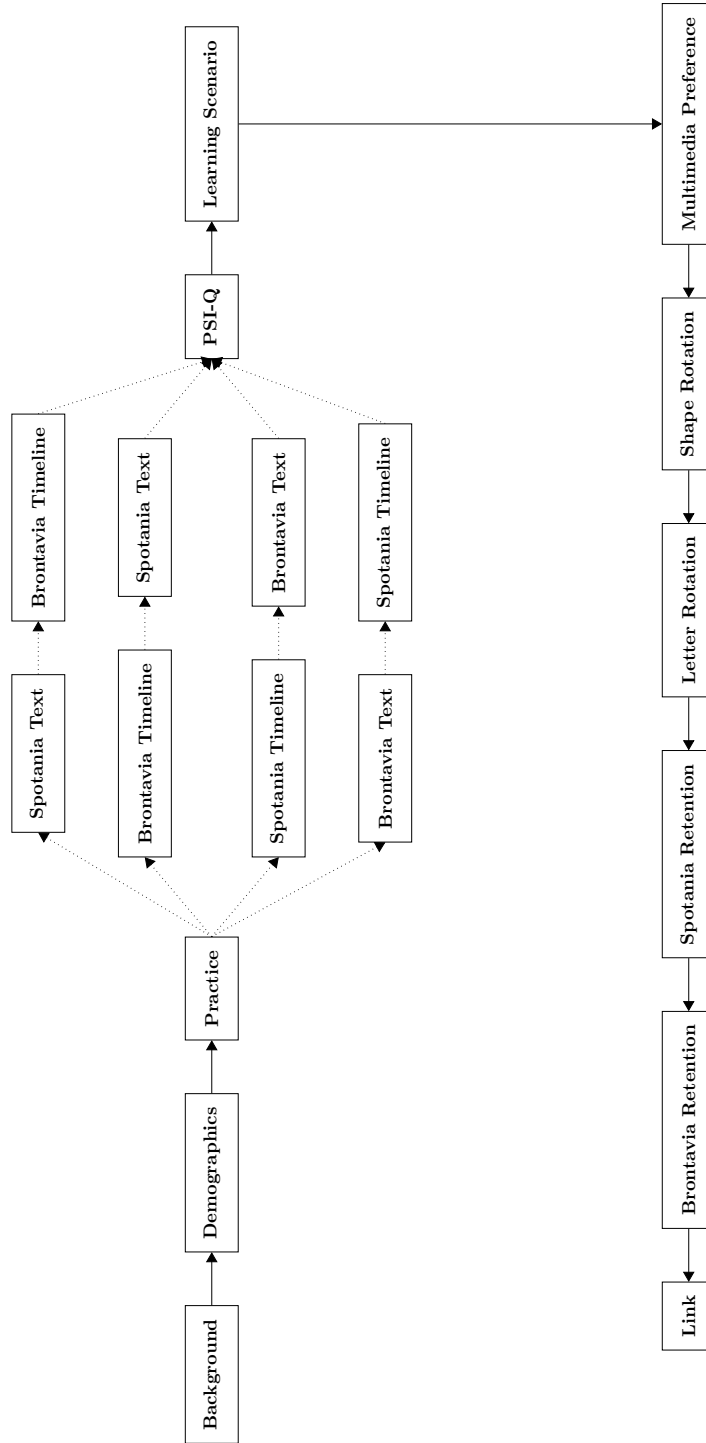
For the first session this study used exactly the same fictional texts, timelines, and comprehension and immediate retention questions as study two.

The second session contained the delayed retention assessment involving 12 questions on each of the stories. These are shown in Tables D.1 on page 262 and D.2 on page 263. The questions were again of three basic types, entry of a number, selection from multiple choice, and placing items in order.

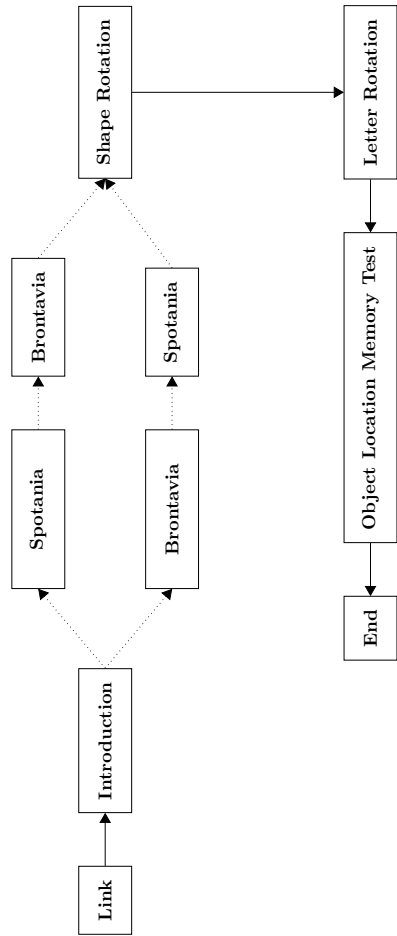
### **5.3.2.4 Mental rotation ability**

**Shape rotation** The underlying materials for the Shepard and Metzler mental rotation task (Shepard & Metzler, 1971), where participants are shown pairs of perspective drawings of three dimensional regular shapes made of blocks, were the same as those used in study two, being obtained from a set published by Tarr (2015). For this study, however, two groups, one for each session, were produced. Each group contained 60 pairs, half of which were congruent. The congruent pairs included equal numbers of each rotation angle of 0, 20, 40, 60, 80, 100, 120, 140, 160, 180 degrees and all pairs were spread randomly through the complete set. The participant was also instructed to use the keyboard (study two had used the mouse).

The two groups of letter rotations were also randomly ordered containing 48 pairs, half of which were congruent, with equal numbers of rotations of 40, 80, 120, 160, 200, 240, 280, and 320 degrees. The same javascript as used in the Shepard and Metzler object rotation tests was incorporated into the pages testing rotations of the letter "F". An example of this test can be seen in Figure D.1 on page 261



**Figure 5.1** – Study Three - Experiment structure - first session



**Figure 5.2** – Study Three - Experiment structure - second session, delayed retention



#### **5.3.2.5 Object location memory test**

This test was based upon the test used by Silverman and Eals (1992). A javascript program was developed for this test and incorporated into the relevant web pages. This displayed an eight by eight grid such as that shown in Figure D.2 on page 264 for 10 seconds, after which the screen was blanked for two seconds before returning with some of the objects changed in location. The participant could indicate by mouse click those cells that they considered had changed. Two such tests were incorporated, in the first test 6 out of 9 items remained static, and in the second, 5 out of 12 remained static.

#### **5.3.2.6 Demographic and Learning Preference questions**

The demographic questions and measures of vividness, learning scenario preference and multimedia preference were all kept the same as in study two.

#### **5.3.2.7 Analysis approach**

A program was developed to score the outputs from the various parts of the study. The methods used are described below:

**Main questions (comprehension and retention)** The questions were scored in the same way as had been used for study two, multiple choice questions being scored with a value of one for the correct answer. Numeric questions were only given a mark if the answer was exactly right, as this is typical of school examinations, which have a prescribed marking scheme. To ensure data were not overlooked, the results were checked to identify any particularly popular wrong answers. The ordering questions received a mark if the items were correctly ordered. The scores were converted to percentages of the maximum score for each phase (12 for comprehension, 6 for immediate retention, and 12 for the delayed retention session 1 week later).

**Vividness** The mean of all the responses to the questions in the PSI-Q subset was used as the outcome, in line with Andrade et al. (2013).

**Learning scenario preference** In line with Mayer and Massa (2003) a score of one was allocated for answers where the visual mode was preferred and the results were totaled to give an overall result in the range zero to five.

**Multimedia preferences** Also in line with Mayer and Massa (2003) a score of one was allocated for each answer where the preference was for the visual help screen and the results were then totaled to give an overall result in the range zero to three.

**Mental rotation** The results of the shape rotation tests in the two sessions were combined together and the angles and response times for correct responses were processed in line with Shepard and Metzler (1971) and Brosnan et al. (2010), discarding incorrect responses and using the slope and intercept of a regression line fitted to the data as the resultant measures. The same approach was used for the mental rotation tests based on the letter F.

**Object location memory** The object location memory score composite measure used by Silverman and Eals (1992) was calculated by summing the number of hits and correct rejections and subtracting the total of false alarms and misses. The results of the two separate tests were averaged to give the final score.

### **5.3.3 Design**

The main element of the first session of the experiment was a "within participant" repeated design with timeline vs text presentation being the independent variable with two levels and two dependent variables (accuracy and response time), both assessed at two distinct time points (while the material is being shown, to gauge comprehension, and after the material has been removed, to gauge immediate retention) The "within participant" variables are the

accuracy and response times. Other "between participant" variables, including age, gender, spatial rotation ability, and multi-media preferences were collected for use in post-hoc grouping analysis.

Fictional materials were used to remove prior knowledge as a direct confound. The experiment was counterbalanced by randomly selecting from four combinations (timeline first vs text first and two different underlying materials) for the comprehension element. During the first session the participant's demographic information was collected, their self rated vividness of visual imagery (Marks, 1973; Andrade et al., 2013), and Multimedia and Learning Scenario Preferences (Mayer & Massa, 2003) were collected, and an assessment made of their spatial ability using a set of pairs of rotated shapes as used by Shepard and Metzler (1971), and a set of pairs of rotated letters as used by Beste et al. (2010). These spatial tests also acted as distractor tasks to prevent any rehearsal of the fictional stories, before the baseline retention questions (Hinze et al., 2013).

The second session consisted of a counterbalanced (in terms of which story was considered first) set of retention questions followed by rotation tests and an object location memory test.

### **5.3.4 Procedure**

**Session One** The study took place in a quiet psychology laboratory on the university campus. After a welcome to the first session by the researcher and allocation to a computer, participants were given time to study the background and to sign the consent form. The questionnaire was viewed via a standard browser (Google Chrome) with the participant able to adjust the window size up to the maximum of the screen resolution of 1366x768 pixels on a 20 inch Liquid Crystal Monitor. The first web pages of the study provided background information and the participant was then asked by the software to enter their demographic information. The next pages provided an explanation of the experiment and incorporated practice examples of three of the basic question types (numeric, selection, and ordering). The main questions followed, with two distinct groups, one of which was timeline based and

the other text based. The first page in each group provided the story in text or timeline form and the participant was allowed to take as long as they liked to study it (they were also told that the information would be repeated on every subsequent page of the comprehension test). For the subsequent pages containing questions, the participant was asked to answer as quickly and accurately as they could (the comprehension assessment).

The next set of questions used self report to assess Vividness of Visual Imagery (Marks, 1973), Learning Scenario, and Multimedia Learning Preference (Massa & Mayer, 2006; Mayer & Massa, 2003). The participant was then presented with pairs of shapes as used by Shepard and Metzler (1971) and asked to identify whether these were the same or different. After this task, which also acted as a distraction, the participant was asked a set of questions on one story and then a set on the other, with no material shown (the immediate retention assessment). The participant also provided a memorable word/letter sequence to form an anonymous link between the first and second sessions.

**Session Two** The delayed retention tests took place a week later ( $7 \pm 2$  days). This period was chosen to match periods used by (Roediger & Butler, 2011) and Cepeda et al. (2008) when assessing retention. The second session took a very similar computer based form to the first session but there was no display of the test material. After entering their participant chosen word/letter sequence to link the data for the two sessions during analysis, a set of questions (delayed retention assessment) was presented covering each of the stories in turn (the story that was asked about first being selected at random). These were followed by further mental rotation tests. Finally two tests of object location memory were performed. After completing the session the participant was thanked for taking part, and provided with a debrief if requested.

## 5.4 Results

### 5.4.1 Group balance

The groups did not differ significantly in terms of their proportion of gender,  $\chi^2(3) = .323$ ,  $p = .96$ , nor in terms of age,  $F(3,44) = .323$ ,  $p = .956$ ,  $\omega^2 = -.003$

### 5.4.2 Participant Accuracy

Examination of answers given to the numerical questions showed no common errors. The six distributions of total marks for the two conditions (text vs timeline material) and the three measurement points (initial comprehension, immediate retention, and delayed retention) were checked for normality using Shapiro-Wilk tests since this test was suggested by both Thode (2002), and Razali and Wah (2011). The detailed results can be seen in Table I.1 on page 279. Although the distribution of marks for the timeline condition in the delayed retention session was not significantly different from normal ( $W = 0.96$ ,  $p = .106$ ), the remaining five distributions were found to be significantly different from a normal distribution ( $W$ 's  $< 0.96$ ,  $p$ 's  $< .05$ ). The distributions showed some skew, with inconsistency of direction preventing use of a transform since the same transform would need to be applied to each in order to be able to perform valid comparisons (Field et al., 2012). There was also a wide difference in variances between the groups suggesting that the Wilcoxon-Mann-Whitney non parametric test would not be appropriate (Wilcox, 2016). The paired comprehension results were therefore compared using Yuen's method as in study two.

**Comprehension** This analysis showed no significant difference between the distribution of marks for the text condition with a 20% trimmed mean ( $M = 11.3$ ), and the corresponding mean for the timeline condition ( $M = 10.9$ ) when assessed using Yuen's method,  $T_y(29) = -1.56$ ,  $p = .129$ , 95%  $CI[-0.92, 0.12]$  with an explanatory measure of effect size ( $\xi = 0.22$ ).

**Retention** For the retention results, since these were measured in two sessions and in both conditions, a repeated measures ANOVA approach would usually be used but the parametric form of this analysis was ruled out by the skew and deviation from normality of the data. For dependent data such as these, the non-parametric Friedman test could have been a suitable way to compare the distributions, however Baguley (2012) considers it to be a relatively inefficient test, while many others recommend the use of robust methods for data with skew, and where normality cannot be assumed (Larson-Hall & Herrington, 2010; Erceg-Hurn & Mirosevich, 2008), with Wilcox (1992) stressing the issues of power and relevance that could be raised if they were not used, and recommending the use of a robust ANOVA using trimmed means. This was the approach selected, and was undertaken by using the R function `rmanova` from the package `WRS2`, based upon Wilcox (2016). The robust analysis of variance showed that there was a significant difference between the distributions of the four measurements,  $F(3, 87) = 56.6, p < .001$ . Performing multiple comparisons with Rom's method for controlling the family wise error (using R function `rmmcp`) as suggested by Wilcox (2016), showed that there was no significant difference in the results for immediate retention, with 20% trimmed mean score for the text condition ( $M = 77.2$ ) and ( $M = 71.7$ ) for the timeline  $\hat{\Psi} = 2.78 (-4.04, 9.59), p = .41$ . For the delayed retention, however, the 20% trimmed mean score for the text condition ( $M = 55.6$ ) indicated a significantly higher level of information retention than that for the timeline condition ( $M = 47.5$ ),  $\hat{\Psi} = 8.33 (2.41, 14.25), p = .002$ .

### 5.4.3 Participant Response Times

The response times for the questions ranged from 1.7s to 114s with  $M = 25.8, SD = 20.3$  and were non-normally distributed, with skewness for the whole set of measurements of 2.17 ( $SE = 0.24$ ) and kurtosis of 6.29 ( $SE = 0.24$ ). The distributions varied between question types and showed evidence of a speed vs accuracy trade off (Anderson, 1981) in the same manner as the response time results of study two. The response time comparisons

were therefore performed the same way as for study two, by being restricted to the cases where the participant had given correct answers in both conditions.

**Comprehension Question Response Times** The distributions were compared using Yuen's method for paired distributions (Wilcox, 2016). Significance values were adjusted using Holm's method (Holm, 1979) and comparisons for each question can be seen in Table I.2 on page 280. For three of the comparisons the differences are significant, and in each case the timeline case has the lower trimmed mean. These cases are reported here: For question three, the 20% trimmed mean time of 11.98s for the timeline condition was faster than the text condition which had a mean of 15.52s

$T_y(25) = 3.16$ ,  $p = .041$ , 95%  $CI[1.23, 5.85]$ , for question seven, the 20% trimmed mean time of 13.19s for the timeline condition was faster than the text condition which had a mean of 18.72s  $T_y(24) = 3.44$ ,  $p = .024$ , 95%  $CI[2.21, 8.85]$ , and for question eight, the 20% trimmed mean time of 12.23s for the timeline condition was faster than the text condition which had a mean of 24.94s  $T_y(26) = 6.32$ ,  $p < .001$ , 95%  $CI[8.58, 16.84]$ . For all other questions time differences were not found to be significantly different.

**Initial Retention Question Response Times** The distributions were compared using Yuen's method for paired distributions. Significance values were adjusted using Holm's method Holm (1979) and comparisons for each question can be seen in Table I.3 on page 280.

In only one of the six questions were the results significant, this was for question one where the 20% trimmed mean time of 5.68s for the text condition was faster than the timeline condition which had a mean of 7.07s,  $T_y(29) = -3.98$ ,  $p = .003$ , 95%  $CI[-2.1, -0.67]$

**Delayed Retention Question Response Times** The restriction to the cases where the participant had given correct answers to questions corresponding to both study conditions combined with the lower accuracy scores overall resulted in too few results to compare for

	<i>Age</i>	<i>MRS</i>	<i>MRI</i>	<i>FRS</i>	<i>FRI</i>	<i>MMP</i>	<i>LSP</i>	<i>CompD</i>
<i>MRI</i>	0.24	0.44**						
<i>FRS</i>	-0.03	-0.03	-0.24					
<i>FRI</i>	0.07	0.34*	0.55*****	-0.73*****				
<i>MMP</i>	0.21	-0.14	0.13	-0.04	-0.10			
<i>LSP</i>	-0.10	-0.12	-0.14	0.20	-0.18	-0.22		
<i>CompD</i>	0.02	0.11	0.17	-0.17	0.24	0.05	-0.06	
<i>RetnD</i>	-0.01	-0.02	0.10	0.04	0.02	0.14	0.10	0.39**

MRS = Metzler Slope, MRI = Metzler Intercept, FRS = F Rotate Slope, FRI = F Rotate Intercept, MMP = Multimedia preference, LSP = Learning Scenario Preference, CompD = Comprehension score diff, RetnD = Retention score diff, \*P < .05, \*\*P < .01, \*\*\*P < .001, \*\*\*\*P < .0001,

**Table 5.2** – Study Three Correlation Table

this test.

#### 5.4.4 Other Variables

The participant preferences in respect of Multimedia Preference (MMP), Multimedia Learning Preference (MLPQ) (Massa & Mayer, 2006) and vividness of visual imagery (Marks, 1973; Andrade et al., 2013) were collected. These, together with the demographic variables of gender and age, score for the object location memory test and slopes and intercepts for the two mental rotation tasks (Metzler\_slope, and Metzler\_intercept and F\_slope and F\_intercept respectively), were examined, together with the differences between participant scores for the text and timeline conditions during comprehension, immediate, and delayed retention. The resulting Pearson correlation matrix can be seen in Figure 5.2. The only significant correlations were between broadly equivalent measures such as the two mental rotation tests and between the comprehension and retention scores (repeated measures of the same participants). There were therefore no meaningful strong correlations.



### 5.4.5 Overall Timing

Overall timings were recorded to assist the design of the subsequent studies. The familiarisation element of the comprehension test was studied for a mean time of 127s ( $SD = 83$ s) with less time taken on average for familiarisation with the second set of material ( $M = 64$ s,  $SD = 55$ s). Text stories were studied for a mean time of 130s ( $SD = 84$ s), and timeline versions for a mean time of 61s ( $SD = 49$ s).

## 5.5 Discussion

This study used the same primary materials, and a broadly similar procedure, to those used in study two, but extended and increased the number of measurements of spatial ability, and included an additional retention measurement a week after the first measurements. The later retention measurement used a counterbalanced design (study two had counterbalanced the comprehension element but used a fixed sequence of questions for the immediate retention phase). Based on the literature (Mayer, 2014a) it was expected that timelines would result in better comprehension (Seufert, 2003), and retention (Schweppe & Rummer, 2012), scores, however studies one and two had provided varying results, with support for hypotheses one and three (concerning accuracy of comprehension, and retention respectively) only found in some cases (study one showed some support for hypothesis one, with timelines improving comprehension in one year group, while study two not only did not support hypothesis one, but it also suggested that the improved retention expected from timeline material in hypothesis three might instead favour the text representations). This study was designed to build and extend upon study two, adding the delayed retention test (Cepeda et al., 2008) after a delay period used by other studies (Roediger & Butler, 2011) and also compatible with school teaching where an initial reinforcement test may well take place a week after the lesson. The study collected, and aggregated, data on spatial ability (mental rotation) (Shepard & Metzler, 1971) measurements during both sessions, improved the

accuracy/reduced the measurement variance by using keyboard rather than mouse responses to the rotation tests, and increased the assessment of spatial ability by incorporating a test concerning object location memory (Silverman & Eals, 1992). The changes addressed the issues identified in study two, and focused on the first three hypotheses of the set, addressing the comprehension and retention aspects of timelines compared to text.

A key change in this study was the increase in spatial measures in an effort to ascertain if spatial ability could be a mediating factor which was masking the effect of the timelines. However, this study again found no meaningful correlation, of either accuracy or response times, with any of the individual differences/preferences measured using the vividness measure PSI-Q (Marks, 1995; Andrade et al., 2013), Learning scenario preference (Mayer & Massa, 2003), and Multimedia preference (Massa & Mayer, 2006), suggesting that, within these studies (two and three) neither spatial ability performance, nor learning/multimedia preferences are useful indicators for whether the learner will benefit from using timelines to learn. The absence of an effect from spatial ability measurements contrasts with the findings of a meta-study by Hoffler (2010), but the effects highlighted there mostly relate to dynamic material involving animations and sound, and to distinctions between high and low ability learners, this study, using mostly university based participants, may not have involved a wide enough range to encounter any effects. A similar observation can be made in respect of the vividness, and learning preferences, there may not have been enough variation present, and similar multimedia studies have also not found a link between such preferences and multimedia learning (Massa & Mayer, 2006) while wider studies have cast doubt on learning preferences in general (Hattie, 2008; Sutton Trust, 2018).

As in study two, no significant differences for accuracy were found during the comprehension testing, which repeated the findings of Korallo (2010b) and Prangmsma (2007). Analysis of the retention data showed a difference from study two with no significant difference during immediate retention, where study two had shown text performance to be better. Interestingly the delayed retention test, which was not used in

study two, did show a significant difference and, as in the retention test of study two, the difference indicated better performance by those in the text condition. This suggests that while the timeline did not hinder the participants at the comprehension stage, information learned using this method was not as well retained for the retention tests. This is contrary to expectations based on the literature where Masterman and Rogers (2002) had reported a positive impact of timelines (albeit based upon teacher feedback rather than any direct, quantitative, measures), and many studies using the CTML (Mayer, 2014a) had shown materials using multimedia to lead to greater levels of learning.

A second aspect of this study was to look at whether timelines had an impact on the cognitive load during comprehension, using response times as a measure (based upon the results of study two which had shown this to be a valid methodology). It was anticipated that timelines would reduce cognitive load as they help to provide comprehensive overviews with connections already highlighted, thus we would expect to see participants responding quicker to questions when the material had been in the timeline condition.

Analysis of the response times as a measure of cognitive load during the comprehension stage indicated that for the three questions where differences were significant participants were faster to respond when they had timeline materials. This result was similar to that of study two where five questions had produced significant comparison results, with the timeline condition proving faster in each case. Also, as in study two, participants spent less time studying the timeline materials during the familiarisation element, suggesting that the timelines aided the participant in obtaining enough information to be ready for the next stage.

When the response times for the initial retention questions were examined, these were also in line with the result from study two (although again, only a single question showed a significant difference) and were also in the opposite direction to that which had been found for comprehension, since it was the text condition that resulted in the shorter response times. It would have been interesting to see if this difference was sustained for the delayed

retention test but unfortunately there were insufficient analysable data due to the need to limit the comparisons to the question pairs where participants got both answers correct. Although not part of the hypothesis being studied here the immediate retention question response time difference found in both studies could be worth exploring in future work. From these results it would appear that the timeline materials proved advantageous at the comprehension stage of assessment as, although they did not improve accuracy scores, they did not hinder scores, and participants spent less time answering the questions, suggesting that the timeline materials made it easier for the learner to access the correct response. However, the assessment results for the delayed retention showing text scores to be higher, suggest that the timeline information was unlikely to have been successfully stored for retrieval. This is further supported by the response times during the immediate retention; again there was no difference in accuracy, but this time the text responses were quicker, indicating that recalling the answer from text based materials required less cognitive load, suggesting it had been successfully learned.

It can be concluded that for study three, timelines were shown to be beneficial for efficiency of comprehension, but disadvantageous in later recall. In contrast, text took longer to comprehend, but led to better accuracy during the delayed retention phase. Situating this in the real world, study three suggests that timelines are useful for quickly extracting key information, but are not beneficial if the information needs to be retained.

It is, however, unclear whether the long term benefit of text for retention, was due to the material being presented in a text form, or if it was due to participants spending longer reviewing the text materials. In order to consider this further the next study will look to control the time spent with the resources.

## **Chapter 6**

# **Study Four: An experimental study examining text versus timelines and self creation of each type**

### **6.1 Chapter Overview**

The results of studies two and three showed that text, rather than timelines, proved to be beneficial during the retention testing, immediately in study two, and delayed for study three. These results were contrary to expectation based on the literature, in particular studies from Masterman and Rogers (2002) who noted strong teacher support for the effectiveness of a timeline based multimedia resource in their experiments, and research in the CTML field, which, although not explicitly looking at timelines, noted improvements in learning when graphical resources are also provided. The rationale behind these inferences is that providing learning materials which invoke both the verbal and visual streams of memory will lead to a greater depth of retention, and ultimately an increase in learning. While studies two and three contradicted the literature, the earlier study one, which was based in a school environment, did show a teacher assessed learning benefit to using timelines.

Therefore this fourth study was designed to explore further this discrepancy between results and literature. Study four was designed to more closely mimic a school environment with the conditional material (timeline or text) provided as supplementary material. This is more typical of an education experience where teachers will use multiple resources to teach a topic. The provision of supplementary materials, either provided for, or generated by, the participant, aligns with the Vygotskian theory of the zone of proximal development, where the learner can use resources to build on existing knowledge (Chaiklin, 2003). In this study each participant was provided with a baseline knowledge on a topic, and the independent variables would vary the type of supplementary material received.

As there has been research on the benefits of students creating resources, and in particular drawing them (Ainsworth, Prain, & Tytler, 2011; Van Meter et al., 2006), this study also included a condition exploring the difference between supplementary materials that are provided or generated by the learner. This introduces the Generative Drawing Principle of the CTML (Schmeck et al., 2014; Schwamborn, Mayer, Thillmann, Leopold, & Leutner, 2010), and it was anticipated that this would lead to a clear distinction between timeline and text materials, through the added value of drawing a multimedia resource (timeline) compared to writing a summary text.

Having noted in previous studies that participants differed in the amount of time spent looking at text vs graphical (timeline) information, the study was developed to examine differences in participant retention of fictional historical information in one of two formats, textual and timeline. In each case additional summary information in the same format was either presented to the participant or generated by the participant. The intention was to explore further the differences in learning from text and learning from timelines found in previous studies and to examine any effects of self creation of material.

## **6.2 Introduction**

### **6.2.1 Rationale**

As discussed in the literature review, the CTML introduces a number of principles which can be used to guide learning with multimedia. While some of these focus upon multimedia facets such as animation and audio presentations, a number of the more relevant principles have been used as tools within the design of studies and their associated materials to maximise learning. This study continues to explore the principles used in previous studies, but introduces the Generative Drawing Principle as another facet which may impact the extent to which timelines can benefit learning. Due to the requirements of the balanced design, and a desire to design an experiment which was similar to a school-style learning experience, this study did not focus on the split attention principle or the redundancy principle. It is considered acceptable to do this as the CTML principles are not mutually exclusive, and it may not always be feasible to employ all of the principles.

The studies in this thesis have investigated a number of aspects of learning from timelines. Study one involved using timelines in a school environment examining the effect on assessment results of providing additional timeline material to a group of students and found a small (but significant) advantage for the use of timelines in this year 8 group. Studies two and three were laboratory based and compared comprehension and retention in terms of both participant accuracy and response times for questions regarding time related material presented in timeline and textual format (both studies involved tests of immediate retention, but study three also included a second retention test after a delay of one week). Within those studies, no significant advantage was found for the use of timelines, and, indeed, in some situations (immediate retention in study two and delayed retention in study three) text versions resulted in significantly higher accuracy scores. However records of the amount of time spent on the initial materials indicated that participant spent longer looking at the original material when it was text, in comparison to timelines. A discussion of this

difference in response times considered whether timelines were proving beneficial as they required less effort to comprehend. To gain further understanding of this aspect, this study is designed to control for the difference by enforcing a limit on the time spent on the initial materials. This also fits with real life examples, where timelines are typically used in a school scenario, where the teacher would be time limited.

Studies two and three were based upon the total material for the participant to study being in one format or the other, so that when studying the story in timeline format, the timeline was the only material available to learn from. Study four returns to a more natural situation, similar to that used in study one within a school, of graphical material being provided as an adjunct to the primary text, to explore how this affects learning. The primary material used by all participants in this study was a text based historical story but some groups used an additional summary text while the others used an additional timeline summary. While this may be considered not to comply with the split attention and redundancy principles, it can be argued that although it is providing a limited amount of information in multiple formats, it is not over elaborate, and by introducing a summary document it utilises the signalling principle which discusses how learning is increased when key information is highlighted. This study also returns to the "school style" approach by exploring the difference between provided and individually created resources. There is a wide range of literature concerning the learning effects of self creation of adjunct material, whether text (Leopold, Sumfleth, & Leutner, 2013; Gil, Braten, & Vidal-Abarca, 2010; Spiergel & Delaney, 2016), or graphics (Fiorella & Mayer, 2016; Van Meter et al., 2006; Leopold & Leutner, 2012; Gobert & Clement, 1999; Zhang & Linn, 2011; Bobek & Tversky, 2016). By including this second variable this study was designed to provide an insight into the hypothesis that "Individual creation of timelines will result in greater retention of information content when compared to provided timelines"

Although a computer based test was used to assess the retention, the materials used in this study were provided in paper form rather than on screen which also matches the approach



used in typical classrooms and study one. This final study is designed to seek answers relating to questions that have arisen during studies two and three in relation to how people use the materials and the time spent learning the information. It is also intended to return slightly from the lab style studies of two and three, into a study which more closely reflects a class room experience (the location where timelines are most often used as a tool for learning). Through designing this study to replicate aspects of a school lesson, with limited time to view materials and provision of main documents and summary documents, the findings should provide better ecological validity and ultimately have more realistic real world implications.

### **6.2.2 Aims**

Based on the earlier discussion of the literature and study results so far, the three hypotheses of this study (of the overall total of five) are (numbered as in 2.5):

Hypothesis 1: Use of timelines will result in increased comprehension of information when compared to text.

Hypothesis 3: Use of timelines will result in greater retention of information when compared to text.

Hypothesis 4: Individual creation of timelines will result in greater retention of information when compared to provided timelines.

## **6.3 Method**

### **6.3.1 Participants and Recruitment**

The study was granted ethical approval from the Department of Psychology ethics committee (ethics reference number:16-315) and participants were recruited through a

Group	n	Female	Age				Timeline use (%)	
			Min	Max	Mean	SD	Sometimes	Frequently
A	18	12	18	65	30	12.6	72	67
B	18	11	19	59	34	12.0	89	83
C	18	11	21	56	33	11.2	61	72
D	20	13	20	58	31	10.8	75	70

**Table 6.1** – Study four - Group summaries

variety of means: through the Psychology Department’s Research participation Scheme (RPS), through general publicity on the university noticeboards, and via opportunistic sampling. All first year psychology undergraduates at the university are expected to participate, through the RPS, in at least five hours of research, split into half-hour blocks, and receive partial course credit for a Psychology Laboratory unit for completion of the five hours. No incentives were provided for any other participants apart from snacks being available during the experiment. All participants were assumed to be fluent in English, this being one of the entry requirements to study at the University.

The mean age of the 74 participants was 32 years ( $SD = 11$ ) with minimum 18 years and maximum 65 years, and 47 were female. Most participants (74%) had used timelines before and 73% used them regularly. The participants were allocated, sequentially by order of arrival at the first session, into four groups (described in Table 6.2) and the corresponding descriptive statistics for these groups are given in Table 6.1. Three participants did not return for the second session, hence the inequality of group sizes.

## 6.3.2 Materials and Measures

### 6.3.2.1 Materials

In order to remove a possible confounding effect of prior knowledge or of research by participants between the sessions, the primary material was produced as a piece of fiction.

The initial underlying sequence of events is based upon an actual battle during the American civil war, but every name, location, and object was changed, the sequence of events was adjusted to fit the narrative/diagram better and some of the details were removed to reduce complexity. Each of the materials described (primary text, summary, and timeline) was printed in 10pt Times Roman font on white A4 paper and laminated.

#### **6.3.2.2 Analysis of text used**

The primary text can be seen in Appendix E on page 265 and contains 548 words (51% of which are unique) with a Flesch Reading Ease score of 52.9 which can be considered fairly difficult to read (Farr et al., 1951) but it should also be noted that automated tools are known to take a simplistic view of text complexity (Schrivver, 1989, 2000), and it was important that the content had a certain level of difficulty to replicate a learning environment. The text has a Flesch-Kincaid Grade Level of 12.2. Both scores were determined using the word processing tool, Microsoft Word, Professional Office Plus. There are 12 distinct characters in the story, 11 locations, and 5 other names/items. There are 14 events with 12 links between them.

#### **6.3.2.3 Additional Summary**

The summary, provided to group A, is a much reduced version of the primary text, highlighting the key points and can be seen in Appendix E on page 267. It was produced in a revision note format, with limited extraneous text, using highlighting for key terms, and included every item covered by the assessment questions. The summary text contains 294 words (52% of which are unique) with a Flesch Reading Ease score of 62.5 which is considered a standard text level by Farr et al. (1951), but again it should also be noted that automated tools take a simplistic view of text complexity (Schrivver, 1989, 2000). The Flesch-Kincaid Grade Level was 8.7.

#### **6.3.2.4 Additional Timeline**

The timeline format summary provided to group C is shown in Figure E.1 on page 270. It includes all of the key points covered in the assessment questions and was produced in a style that might have been produced by a student for revision.

#### **6.3.2.5 Equipment for participant creation of additional material**

Note Creation: Participants in group B were asked to create their own summary notes and were provided with blank sheets of A4 sized white paper, a ball point pen, and a pencil for this.

Timeline Creation: Participants in group D, who were asked to create their own summary in timeline format, were also provided with blank sheets of A4 sized white paper, a ball point pen, and a pencil but received, in addition, a set of 20 coloured pens and a 30cm ruler to use as they wished.

#### **6.3.2.6 Computer test questions**

All participants completed a computer based test after studying the materials, and either reviewing or generating the additional materials. In the first (immediate retention) session this started with entry of their participant number and followed with the assessment questions. These are shown in Table E.1 on page 269. The same set of assessment questions were used in the delayed retention test (in studies two and three only a small subset of questions had been identical).

The questions were refined based upon experiences gained from studies two and three where low accuracy (generally much less than a quarter of participants getting answers correct) was found for numeric questions, (especially low in the retention assessments). Selection questions (one of two, three, or four choices), and order based questions (placing three or four items in chronological order), provided results closer to a 50% average. The numeric questions in previous studies also necessitated the participant changing from use of the

Group	Activity	Mode	Delivery
A	Given a text summary	Text	Presented
B	Asked to produce a text summary	Text	Created
C	Given a timeline summary	Timeline	Presented
D	Asked to produce a timeline summary	Timeline	Created

**Table 6.2** – Study four - Group materials

mouse to keyboard and resulted in a much larger variance in associated response times. The questions used in study four were therefore only selection or ordering types.

### 6.3.3 Design

The experiment was a four way mixed design (split-plot) with the type of support material (text summary vs timeline summary) and its method of production (supplied vs created) as 'between participants' independent variables (the four conditions are shown in Table 6.2), and two dependent variables (accuracy and response time), which were assessed by computer questionnaire at two measurement times (during the first session immediately after learning had taken place, with only the collection of demographic variables and example questions in between the learning and the test, and during a second session a week later). The 'within participant' variables were the accuracy and the session. Other 'between participants' variables including age, gender, and multi-media preferences were collected for use in post-hoc grouping analysis.

Participants were allocated at random (sequential allocation on arrival) to one of the four experimental conditions. During the first session each participant was given the same underlying textual primary material (the story shown in section E on page 265), but additional material was provided or generated, according to group, as shown in Table 6.2. The table also shows the mode and delivery categories used in the later analysis. The additional text summary can be seen in section E on page 267, with the timeline version in

Figure E.1 on page 270. The participant was given ten minutes to study the material and asked to memorise as much of the information as possible (names, locations, times, sequences of events).

Inclusion of a controlled time was introduced to avoid inter-participant variability in time spent on text vs timeline seen in studies two and three.

The period of ten minutes was selected by considering that the primary material contained 548 words, with the aim of allowing the average participant enough time to read the material at least twice (Spichtig et al., 2016) and for participants in group B and D to generate their additional materials. Furthermore, informal conversation with the teachers from study one had indicated that teachers are limited in the amount of teaching time they have, and that often tasks are limited to around 10 minutes. If the allocated task time is too short the students will not be able to complete the work, but if it is too long some students may finish and become bored. After completing the study, and before starting the assessment task, all materials were removed.

A second, identical, retention test took place approximately one week later, chosen to match periods used by others when assessing retention (Roediger & Butler, 2011; Cepeda et al., 2008). The reasons for incorporation of the immediate test of retention were two-fold: first, it provided a baseline against which the level of retention in the second session could be compared, and, second, since feedback was given regarding incorrect answers, it should have improved the initial learning level due to the testing effect (Roediger & Butler, 2011; Carpenter et al., 2009; Roediger & Karpicke, 2006). This style of testing again mimics a school environment closely, where students would be familiar with producing a piece of work, having it assessed, and then later being tested on the topic.

#### **6.3.4 Procedure**

The participant was allocated to one of the four experimental groups in order of arrival at the first session (the recruitment process being considered sufficiently random in sign up and

agreement of session date/time). They were also allocated a participant number at this point. The experimental studies took place in a quiet, distraction free, environment. The materials were provided in paper form (laminated for ease of use and retention by the researcher). During the first session the researcher welcomed the participant, provided them with the appropriate set of instructions, and with consent forms (a copy for the researcher and a copy for the participant to retain).

After reading the instructions and signing the consent forms the participant was given 10 minutes to study the main material and either review the summary document (text or timeline) or produce a summary document (text or timeline). Following the 10 minutes all of the materials (including any self created study aids) were retained by the researcher. The participant then completed a questionnaire collecting age, gender, experience with timelines, and graphical/multimedia preferences. The immediate assessment task then began, with questions based on the story content, accuracy feedback being provided on screen at this point. The participant was provided with a reference number to link the two sessions

The second session took place approximately one week ( $7 \pm 2$  days) later. This period was chosen to match periods used by similar studies (Roediger & Butler, 2011; Carpenter et al., 2009; Roediger & Karpicke, 2006) and to match the delayed time used in study three. After reading the instructions, the participant performed the computer based test of their retention, with the exact same questions being asked in this second assessment stage, but this time feedback was not provided, mimicking a more "exam style" experience.

The participant was thanked for their time and offered a full debrief of the experiment.

## 6.4 Results

The groups did not differ significantly in terms of their proportion of gender,  $\chi^2(3) = .186$ ,  $p = .98$ , nor in terms of age,  $F(3, 70) = .314$ ,  $p = .815$ ,  $\omega^2 = -.04$ . The marks for the two

sessions for each of the groups were examined for extreme outliers, but there was none outside three times the interquartile range, and the scores were normally distributed, as assessed by Shapiro-Wilk tests  $W's(17) > .92$ ,  $p's > .13$ . The data were therefore analysed using a mixed ANOVA (BBW) and the residual data were also found to be normally distributed  $W's(38) > .94$ ,  $p's > .09$ .

There was a significant difference for session,

$F(1, 70) = 7.194$ ,  $p = .009$ ,  $partial \eta^2 = .093$  with participants, regardless of grouping, having a lower mean during session two ( $M=11.98$ ), than session one ( $M=12.63$ ).

There was a significant difference for delivery,

$F(1, 70) = 8.144$ ,  $p = .006$ ,  $partial \eta^2 = .104$  with participants who were provided with materials having a higher mean ( $M=12.93$ ), than those who produced their own ( $M=11.68$ ).

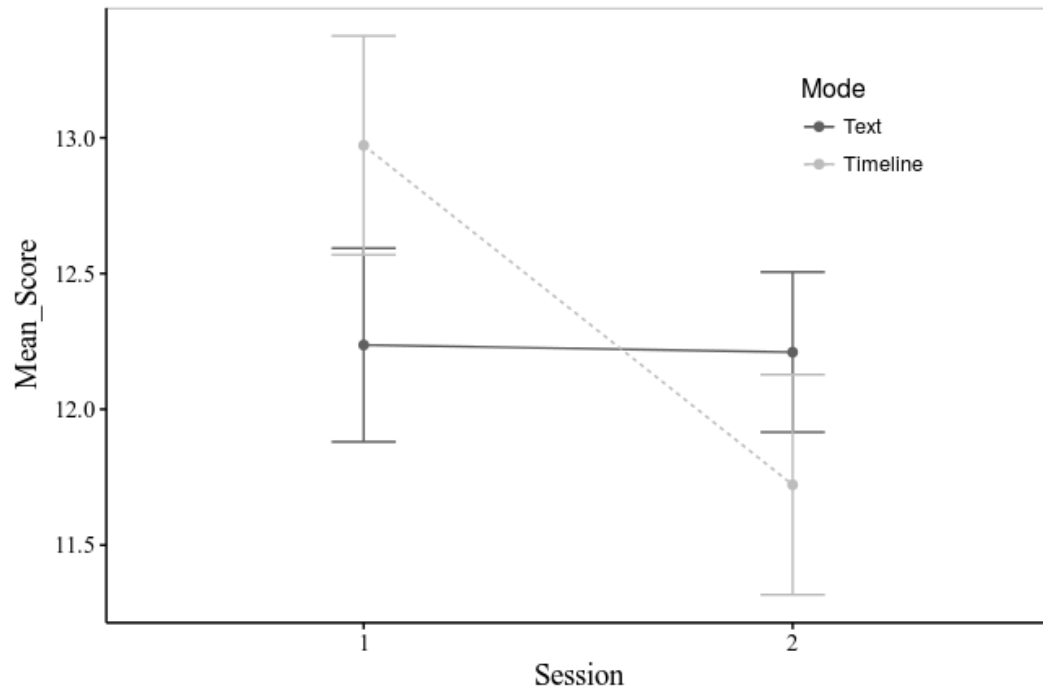
There was no significant difference for material,

$F(1, 70) = 0.038$ ,  $p = .845$ ,  $partial \eta^2 = .001$ ; timeline condition ( $M=12.26$ ), text condition ( $M=12.35$ ).

There was no statistically significant three-way interaction between session, delivery and material,  $F(1, 70) = 0.022$ ,  $p = .881$ ,  $partial \eta^2 = .000$ . There was a statistically significant two-way interaction between session and material,  $F(1, 70) = 6.24$ ,  $p = .015$ . All other two-way interactions were not statistically significant (between session and delivery,  $F(1, 70) = 5.347$ ,  $p = .119$ , between delivery and material,  $F(1, 70) = 0.149$ ,  $p = .701$ ). A simple main effects analysis was performed to follow up on the significant two way interaction between session and material. There was no significant difference between materials at either session one,  $F(1, 70) = 1.836$ ,  $p = .180$  or at session two,  $F(1, 70) = 1.097$ ,  $p = .299$ .

The results indicated that the effect of material was significantly different for sessions one and two, with the text mode showing a large decrease in mean score at session two (Text Session one,  $M = 12.97$ , Text Session two,  $M = 11.72$ ), while the timeline mode only showed a small decrease in mean score in session two (Timeline Session one,  $M = 12.28$





**Figure 6.1** – Study four - interaction between mode and session

compared to Timeline Session two,  $M = 12.24$ ).

In addition the significant difference in the main effect for delivery, but with no significant difference found in the mode-delivery interaction, suggests that while being provided with the materials led to higher scores, this was not mediated by the material (text or timeline).

## 6.5 Discussion

The key aim of this final study was to return to a more classroom style experiment, closely replicating the type of learning scenario that would occur in the real world. Continuing to be guided by the CTML and its associated principles, this study sought to explore how summary type (text vs timeline) could affect the learner in their immediate and delayed retention. By controlling for the time spent on the resources, and providing all the participants with the same novel information, this study was able to explore learning in a laboratory environment. The inclusion of an additional variable, exploring provided vs individually generated summary documents, allowed exploration of the generative drawing

principle of the CTML (Schmeck et al., 2014; Schwaborn et al., 2010). A significant effect was seen for the delivery mode, provided compared to generated, though this effect was in the opposite direction to that suggested by the principle, with participants who experienced the provided materials scoring higher on average across the two assessment sessions. As with many other CTML related studies, however, it could be argued that the underlying material, a chemical process in the case of Schmeck et al. (2014), and a biological process in the case of Schwaborn et al. (2010) are much more to do with gaining a mental model of the processes themselves than just the elements/facts involved. Such material is perhaps more dynamic in nature than a timeline even with a direction and lines of flow. Furthermore, this effect was not influenced by the material type, with no significant interaction between material and delivery. This result did not support hypothesis 4, "Individual creation of timelines will result in greater retention of information when compared to provided timelines". This result was unexpected and may have been due to the provided resources being too closely linked to the assessment questions. Future research could explore whether additional guidance about the key aspects along with self generation leads to an improvement in scores. Considering the developmental literature, self generation of materials should be beneficial as it allows the learner to take control, and form their own links. However, as Piaget and Vygotsky both noted, learners need to be guided, either by peers or their teacher, to ensure that their learning is accurate and relevant. In this instance it may be that those in the self generated condition did not have enough support and guidance about what to include in their generated materials. While disappointing for this study since the results were not as expected, it is an important issue to raise in terms of education. Frequently schools encourage students to make their own notes from resources and to summarise information, but this study has highlighted that, while this is an important skill, it is perhaps more important that the student receives relevant guidance to develop their comprehension and summarising skills.

The significant difference in session suggested that regardless of material or delivery, the

participants were not fully "learning" the material and storing it effectively for retrieval at the delayed point. The absence of any difference here is in line with other research (Korallo, 2010b; Prangmsma, 2007) that also showed the learning effect was no longer evident after a delay. On reflection it may have been helpful to have included a comprehension task in this study, as it is not clear if the students were ever able to extract all of the information from the materials .

One result which warrants further investigation in future research is the interaction between material and session, since the present results indicated that performance in sessions one and two was influenced by the material (timeline or text). In particular, while both showed a decrease in accuracy in session two, this was more extreme in the text condition. It could be suggested from this, that while neither timeline nor text was successfully supporting the participant fully in their learning, the timeline condition showed less deterioration with time. It is possible that a further retention test may have established whether the downward trend for both text and timeline would have continued, or whether the timeline condition would have plateaued, indicating a limited long term benefit from using timeline materials. As the timeline condition showed better consistency over time, it could be said that it did help in the learning process as loss during delayed retention was minimised, which was possibly due to the timeline providing a better oversight of the historical events, and providing additional encoding of the learned information through the visual and verbal routes.

It is clear from this study and the previous studies that learning is not an easy concept to measure, and furthermore that it has not been successfully shown that timelines can measurably improve student learning, either at the comprehension stage or during retention. However it has provided additional insight into this field of research, and the next discussion chapter will look to continue this reflection and consider what the complete set of studies have shown.

# **Chapter 7**

## **Overall Discussion**

### **7.1 Chapter overview**

This chapter discusses the outcomes of the studies described in the thesis, drawing out common themes, and relating them to the wider literature. It reflects upon study design decisions, and on reasons why some results were not as predicted by the hypotheses. The chapter concludes with a review of the contributions made by the thesis both theoretical and practical, and suggests areas for future work.

### **7.2 Introduction**

The links between time and space in the mind are strong, as evidenced by the many spatial metaphors for time found in language and by measurable interactions between time and space, such as that seen in the STEARC effect (Ishihara et al., 2008) and the associated notion of mental timelines (Weger & Pratt, 2008; Flumini & Santiago, 2013; Bonato et al., 2012; Bender & Beller, 2014; Aguirre & Santiago, 2017). Research literature reviewed in chapter two highlighted these links and described how, in efforts to convey the flow of time, and relationships between events in time, there has been a long history of making these time and space linkages concrete in the form of timelines, usually as labeled events on a diagram

with their spatial position reflecting their position in time and with interconnecting lines to show relationships. There are other, more complex, approaches, such as those based upon virtual reality (Korallo, 2010b; Korallo et al., 2012), but two dimensional diagrammatic representations are the most likely representations to be thought of when discussing the role of timelines. This is certainly the case when their use is strongly promoted for the teaching of history by some in education, and its associated politics. The literature review in chapter two described some of the strong opinions supporting the use of timelines as a way of conveying an understanding of chronology and of the "Big Picture", needed for a good understanding of history. The lack of scientific evidence supporting these opinions, however, was identified as a gap in knowledge, both of the general, intrinsic value of timelines, and of their use in real world education. This gap is something the research presented here has aimed to reduce. Examples exploring virtual reality based timeline use, investigating learning and collaboration effects around their production, and scaffolding to support primary school age understanding of chronology through the use of multimedia (Masterman & Rogers, 2002), had contributed to knowledge in this area, but the gap remained large, with none of the identified timeline research utilising "within participant" designs. This was an improvement suggested by Korallo (2010b), and used by Voss and Silfies (1996) when comparing types of text in history learning. The inclusion of participant generation of materials had also not been explored, for example the work by Prangma, Boxtel, and Kanselaar (2008), involved, collaborative, but not individual, construction of timelines. Thus this thesis also sought to add to the literature by conducting two of the studies with within participant methodologies, and through the final study which started to explore differences relating to self generation of materials, something which research into student learning and self generation suggests may be beneficial (Lin et al., 2017; Schwaborn et al., 2010). In order to situate and interpret this research, a number of theories with relevance to learning from timelines were described in chapter two, with Mayer's (2005) Cognitive Theory of Multimedia Learning (CTML) being one of the most common and comprehensive.

Investigations of its applicability outside of the science and mathematical fields were identified as being limited (Tabbers, Martens, & Merrienboer, 2004). The associated research also tended to use somewhat limited materials and evaluation activities, usually through an immediate test rather than any delayed test/longitudinal approach (Schweppe & Rummer, 2012). Thus it was anticipated that again the work of this thesis could add to the existing body of research, filling in holes and extending current understanding.

The CTML contains many design principles, but the need for the material and its delivery to match the optimum cognitive capacity for the learner is an underlying requirement for all.

This fundamental foundation is based upon Sweller's Cognitive Load Theory (CLT) (Sweller, 1988, 1994, 2010; Sweller & Chandler, 1994; van Merrienboer & Sweller, 2005; Paas et al., 2004), and the notion that learning will not occur if the material or approach induces a cognitive load that exceeds the learner's capacity at that point. To assess the effects of different materials etc., it is therefore necessary to be able to compare the levels of cognitive load that they induce. Of the variety of ways used to assess cognitive load, the CTML based research discussed in chapter two mostly made use of a self report instrument considered to be effective in the field (Paas, Van Merrienboer, & Adam, 1994).

Other approaches, such as dual tasks, and secondary physiological measures, can be intrusive and reduce the ecological validity of the experiments. One intrinsic, and measurable, aspect of a participant's activity during an experiment, however, is their response time to questions. There can clearly be some potential confounds in comparisons where question types and their expected responses differ, but this was also considered to be a gap in research requiring exploration in connection with learning from timelines.

Therefore a key aspect of study two was to establish the extent to which reaction times might be comparable with self reported load.

In summary therefore, the research gaps that this thesis set out to explore were, 1) The differences between learning from timelines and learning from equivalent text, 2) The effect on learning of self creation of timeline materials, and 3) A comparison between response

time measurements as indications of cognitive load and cognitive load assessed through self report measures.

The aspects of learning were split into initial comprehension (which is essential for retention to occur) and retention. Assessment of "transfer" as a measure of learning used by some of the other studies in the area of CTML was excluded to allow the focus to be upon a solid, objectively tested, foundation which could then be extended to transfer by others in later studies. Unlike many of the other studies, however, retention was assessed (in some of the experiments) both immediately and after a delay (Schweppe & Rummer, 2012), as the latter was considered to be more representative of the educational use of timelines.

The research gaps described above led to the definition of five hypotheses:-

Hypothesis 1: Use of timelines will result in increased comprehension of information when compared to text.

Hypothesis 2: Use of timelines will result in shorter question response times during comprehension of information when compared to text.

Hypothesis 3: Use of timelines will result in greater retention of information when compared to text.

Hypothesis 4: Individual creation of timelines will result in greater retention of information when compared to provided timelines.

Hypothesis 5: Question response times will have a positive correlation with self reported estimates of cognitive load.

To address these hypotheses and contribute towards closing the underlying gaps, four experimental studies were developed. In the next section these four studies are summarised and followed by a discussion of the results from all studies.

## 7.3 Summary of the research

### **Study one - A study of using a timeline in a UK secondary school**

The aim of study one, described in chapter three, was to explore three examples of timeline use in the educational environment of an English secondary school. A strong focus was placed upon maintaining ecological validity, with the use of teaching staff for the development, delivery in lessons, and subsequent assessment of learning. Validity was also maintained by making use of the existing split school, set based structure (Ross & Morrison, 1989; Black & Wiliam, 1998). Timelines were added to the materials used by one set in each of three year groups (covering Key Stage 3, ages 11-14) and their results, assessed by the teachers, were compared by the researcher in the way that would normally be done by the school. Lack of data for one year group (due to the unexpected departure of the teacher involved), coupled with a significant imbalance between groups in another (a concomitant risk of the use of existing groups for ecological soundness), led to the main study conclusions being based upon only one year group. The results from that group did, however, align with the expectations of hypothesis one (and hence the CTML) that it had been designed to explore, and supporting similar work (Mayer, 2009c; Mayer & Sims, 1994). The group that was provided with additional timeline material did have a higher level of mean improvement against expected level than that achieved by the corresponding control set. This result had, therefore, been demonstrated, albeit in only one study, with only one timeline example, in an approach with a stronger claim to ecological validity than most other investigations in the area.

The preservation of ecological soundness in the design of study one led to a number of research consequences. The groups were pre-existing and could not be chosen for balance (resulting in exclusion of one set of data from the final analysis), and the researcher was not present while the material was being delivered and used and hence could not be certain that the conditions (physical, behavioral, and presentational) between the groups were identical.



Since the material had to be completely relevant to the curriculum, some, but not all, of the participants may have had relevant prior knowledge of the topic, resulting in a potential confound that could not be controlled. The assessments of learning were also made by the class teachers and hence were subjective. Finally, again for ecological validity, the comparisons made were "between" participant groups (with no demographic data other than gender available with which to compare the balance of the groups).

### **Study two - Comprehension and immediate retention**

In the design of study two, the comparison between text and timeline versions of information was made a within participant element, fictional material was used in order to remove any effect of prior knowledge, and the results were objectively measured via automated scoring. The experiment examined the comprehension, and immediate retention level, of time related material in the form of a short, historical, story. The comprehension was assessed while the material remained available for the participant to see, in the same way as for study one, but retention tests took place without the material, and also after other, unrelated, testing activities had been undertaken between the comprehension and retention tests, to ensure that the information was unlikely to be retained by rehearsal in short term memory. The results showed no significant difference in participant accuracy at the comprehension stage but found an accuracy difference when assessing retention, with the trimmed mean accuracy score for the textual format being higher than that for the timeline condition. This was, in the opposite direction to that expected from the CTML, from Hypothesis three, and from study one. Study two also incorporated assessments of participant spatial cognitive ability as measured by mental rotation tests, their vividness of imagery, a range of multimedia related learning preferences, timeline familiarity, occupation and education level, age, and gender. When these were examined for significant interactions none was found.

To investigate response times as a measure of cognitive load, and to understand if timeline material may lead to quicker comprehension response times, participant response times for

every question were collected. Initial analysis of these times showed that their distributions varied widely, both between question types and between correct and incorrect answers. Comparisons between response times could therefore only be meaningfully made for related pairs or identical subsets of questions. In every case where those comparisons achieved significance however (five questions in the comprehension section of the experiment), the trimmed mean time for the timeline condition was lower than that for the text condition hence supporting hypothesis two, at least in the specific examples studied.

In support of hypothesis five the self assessed cognitive load results were compared with the response time measurements and, for the questions where the correlation achieved significance, all the relationships were in the same positive direction, with either weak or medium strength. Conversely, no significant correlations were found between any of the individual preferences, demographic variables, session results or response times. Nor were any significant differences found when grouping the primary results using these preferences and differences (in line with Hoffler (2010) who only found differences where animation material was used).

The overall conclusion from this study was that, whilst there was an indication that some response times may be shorter in the timeline condition during comprehension, and that the initial study time was also shorter in that condition, an accuracy difference was only seen for the retention phase and that this was in the opposite direction to that expected. There was, therefore, some evidence in line with hypothesis two, no evidence to support hypothesis one, evidence for the opposite direction for hypothesis three, and some support for hypothesis five.

### **Study three - Comprehension, immediate, and delayed retention**

Study three was designed to explore the retention aspect further by incorporating a delayed retention test, and to increase the breadth and depth of spatial ability assessment via the inclusion of additional tests of mental rotation and object location memory. The study built

upon the framework used in study two, using exactly the same timeline and text materials for its first session. A greater number of mental rotation tests using shapes were incorporated, together with mental rotation tests using a letter, but otherwise the first session was identical to that of study two. The second session was also broadly similar in approach (although without any comprehension element) and took place a week later. Since the recruitment method was slightly different (mostly undergraduates at the university) the mean age was lower for this study, but results were expected to be comparable. Again, as in study two, no significant difference was found for accuracy in the comprehension session. The immediate retention session, which had shown significantly higher accuracy results for the text condition in study two, did not show a significant difference, although the result approached significance and, as with study two, the trend was in the opposite direction to that expected by hypothesis three. The addition of the delayed retention tests allowed the two sets of retention results (immediate and delayed) to be analysed as a repeated measures robust ANOVA equivalent and this did show that the retention of text based material resulted in significantly higher accuracy during the delayed session. Again this was the opposite result to that predicted by hypothesis three.

The response time measurement analysis during comprehension reinforced the support for hypothesis two that had been found in study two. In all cases where the comparisons were significant (3 of 12) the response times in the timeline condition were shorter than those of the text condition. Individual differences including gender, age, highest education level, occupation, and familiarity with timelines, also showed no significant correlations with either accuracy or cognitive load, nor did assessments of spatial ability. Thus these results indicated that while timelines seemingly did not help with learning in the long term (scoring lower in the retention aspects) the participants were able to respond quicker (and maintain accuracy) during the comprehension aspect, indicating that the information was easy to decipher, but maybe the ease prevented the appropriate level of information encoding to form a sustained memory/understanding (Einstein, McDaniel, Owen, & Cote, 1990;

Schweppe & Rummer, 2016).

#### **Study four - An experimental study examining text versus timelines and self creation of each type**

The final study was designed to address hypothesis four concerning the effects of self creation of timeline materials, in line with the generative drawing principle of the CTML (Schmeck et al., 2014; Schwamborn et al., 2010), as well as to continue to examine retention under hypothesis three using a different set of materials and approach. This study returned to a "between" participant group design and also incorporated feedback during the first retention session to increase learning through the testing effect (Roediger & Karpicke, 2006). A further difference was that the study used paper based presentation of materials and a bounded time for their assimilation, both of which made the conditions closer to those that would generally occur in a school, as in study one. A two part test was involved, in a similar manner to study three, with the delayed retention test around one week after the first retention test. The four groups in the study all received a common basic text, which was fictional to remove any confounding effect of prior knowledge, but, in this study, the effect of two types of additional materials (text and timeline) and two modes of delivery of the additional materials (presented vs self created) were explored. A significant primary effect was found for session, all of the participants showing a decrease in score for the delayed retention as would be expected due to the forgetting effect (Ebbinghaus, 1885). There was also a primary effect for delivery with participants who received additional summary material (regardless of the type) performing better on average across the two retention times, in comparison to those who self generated. This would not be expected from other studies of self generation (Schwamborn et al., 2010; Schmeck et al., 2014; Leopold et al., 2013) although it may have been due to the allocated time period being too short to create and study the materials. A significant interaction effect was found between the mode and the session, the two material modes showing different magnitude of trends between sessions,

the timeline condition showed a smaller decrease in scores from session one to two, while the text condition showed a larger reduction in accuracy at time two.

### **General discussion of study results**

The results of the four studies and their levels of support for the five hypotheses are somewhat mixed and are summarised in Table 7.1. Hypothesis one: "Use of timelines will result in increased comprehension of information when compared to text", was supported by study one, but neither study two, nor three, which also examined comprehension, provided any significant evidence either way. To examine hypothesis two: "Use of timelines will result in shorter question response times during comprehension of information when compared to text", studies two and three made response time measurements during the comprehension questions and found evidence to support the hypothesis. When examining hypothesis three: "Use of timelines will result in greater retention of information when compared to text", however, the same studies (two and three) actually found the opposite to apply in those circumstances where a significant difference was found (for immediate retention for study two, and for delayed retention for study three). In those studies it was the text condition that resulted in greater retention. Study four, which had also been expected to support the hypothesis, provided no significant evidence (although the analysis was close to significance for the contrary view to that of the hypothesis). Study four was also designed to explore hypothesis four: "Individual creation of timelines will result in greater retention of information when compared to provided timelines". The study results were limited in their ability to answer this hypothesis as there was no interaction between mode and delivery, preventing the discussion of timelines by themselves. However, comparing the delivery regardless of material mode, indicated that, in this instance, the creation of materials by the participant had not led to an increase in information retention, and in fact those in this condition did significantly worse than those who were provided with materials. Finally, for hypothesis five: "Question response times will have a positive correlation with self reported

estimates of cognitive load", study two did provide some significant evidence of correlation, in terms of direction, between the two measures.

The key points to discuss from the summary above, are that, although supported by the first study, studies two and three did not support hypothesis one, that studies two and three found the opposite effect from that suggested by hypothesis three, and that no evidence was found to support hypothesis four. Hypothesis two was supported in study two, and hypothesis five was supported in study three. The next section discusses each of the hypotheses in greater depth, reflecting on the study results and the literature.

### **General discussion of study materials**

Since the initial research context concerned the support for the use of timelines in history teaching, all four studies used realistically complex materials of the type that could be used for such teaching. The first, school based, study used actual curriculum material, the other studies used fictional material but each included a realistic number of events, linkages, and other detail. Even the timeline that might be considered the simplest, that used for the year 7 group in study one (Figure B.1 on page 238), included the additional complexity of a more conventional, graph like, indication of authoritarianism on the Y axis against time on the X axis and so cannot be considered a simple linear timeline. Table 7.1 summarises the five styles of timeline used. Later discussions reflect upon the extent to which these varied, but complex, timelines might engage with notions of mental timelines, maps, and general multimedia.

Study	Linear	Time marked	Linked detail	Multiple timelines	Complexity	Images	CTML Principles
1 - Y7	No	Only by reigns	yes panels containing cartoons	No	single line, 4 events but also significance to the vertical dimension	Yes (plus timeline itself)	CTML, Split Attention, Spatial Contiguity, Temporal Contiguity, Coherence, Signalling
1 - Y8	No plus 3 line zig zag	Dates but dates mixed with centuries	Yes	No	12	Yes	CTML, Split Attention, Spatial Contiguity, Temporal Contiguity, Coherence, Signalling
1 - Y9	Yes horizontally but not vertically	Years horizontally, specific months vertically	Yes (vertical lists)	vertical detail	8 primary dates, 19 in verticals	No	CTML, Signalling
Study 2 & 3	Yes but multiple lines	Yes, with scale and detail	Text directly on chart (but note colour coding as well)	Yes (4/5)	12/17 events (story dependent)	No	CTML, Split Attention, Spatial Contiguity, Temporal Contiguity, Coherence, Signalling
Study 4	Yes but multiple lines	Yes, with scale	Yes	two primary plus 5 showing periods	17 events	No	CTML, Split Attention, Spatial Contiguity, Temporal Contiguity, Coherence, Signalling

**Table 7.1** – Summary of Characteristics of Timelines used in the studies

## 7.4 Reflecting on the hypotheses

### **Hypothesis One: Use of timelines will result in increased comprehension of information when compared to text**

Based on the previous research, particularly the CTML (Mayer, 2014a; Mayer & Gallini, 1990), hypothesis one was considered in studies one, two and three, though support for this hypothesis came only from study one. The teachers involved in study one both considered timelines to be useful tools for teaching history (much like many of the educationalists cited in the literature study in chapter two), and it was reassuring for them to find that the data that they helped to collect supported this view. Neither of the other two studies involving the assessment of comprehension provided any significant evidence to support this, however. A major difference between study one and studies two and three concerns how the data were assessed. In study one the students' comprehension was defined by the teacher and their subjective assessment of a piece of work, albeit informed by experience, training and curriculum guidelines. In contrast studies two and three employed a more exam style assessment approach with comprehension being graded by an accuracy mark. As discussed in chapter two, the form of comprehension assessed in studies two and three is a very basic form of comprehension essentially examining how easy the participant finds it to look up the information presented in front of them at the time. Very few mistakes were made in either timeline or text mode (more than 80% of participants were within two marks of the maximum in study three for example) indicating that there may have been a ceiling effect which could have limited the study.

In summary, although this hypothesis was supported in a real world, ecologically valid, study, it was not possible to repeat this result in the lab environment. It would be beneficial to explore this hypothesis further and establish whether it was due to the ceiling effect of the scores, the differences in the method of assessment of 'comprehension' between study one and that used in studies two and three, or something else. One possibility is that the timeline



presentations may not benefit from sufficient deeper level processing (Craik & Lockhart, 1972), they may be perceived simply as a superficial image and the information conveyed by the structure may be ignored. Also, only a few questions in studies two, three, and four, related to the sequencing of events where memory of a timeline might be expected to benefit and it is possible that the examination of the timelines during the comprehension phase (in studies two and three) did not induce the appropriate processing level to match the subsequent retention tests (Morris et al., 1977; Scevak & Moore, 1998). The support of hypothesis two, concerning response times during comprehension, by the results of the second and third studies may also be considered to add further weight to the argument here, that timelines do help during comprehension, reflected either through final accuracy scores, or through the speed of understanding the material.

**Hypothesis Two: Use of timelines will result in shorter question response times during comprehension of information when compared to text**

Hypothesis two was explored in study two, and the results indicated that there was support for this hypothesis, although it was limited to five out of the 12 questions where significant differences were found. For those five questions, in each case the mean response times for the timeline questions were smaller than those for the text questions. From this, and the results indicating no difference in comprehension accuracy scores, it could be deduced that information presented in the timeline format could be comprehended to the same level as the text format, but in a much shorter time frame. This is a really interesting result as it suggests that while timelines may not lead to an increase in accuracy, the participants reached the same performance level in a shorter time. This has real world implications in schools in particular where time is a scarce commodity, and the ability to teach students in a shorter time frame is hugely beneficial. However, it is important to note that the performance difference was limited to the comprehension aspect of the study; during retention testing the accuracy for both groups decreased and, in particular, the accuracy score was significantly

worse for the groups with the timeline condition during the retention stage. The key question to consider here, is whether this is due to the text condition leading to a deeper level of learning (Loaiza et al., 2011; Craik & Lockhart, 1972; Morris et al., 1977), or if this difference in retention is an artefact of the time spent during the comprehension phase.

### **Hypothesis Three: Use of timelines will result in greater retention of information when compared to text**

Studies two, three and four all considered hypothesis three, the results of studies two and three both showing significant differences between text and timeline, but in the opposite direction to that predicted by the hypothesis based on the literature. Study four did not show a significant difference between text and timelines during retention. Interestingly, while both studies two and three found significant differences in retention, study two showed this at immediate retention, but this was not evident at the same point in study three, with the difference only showing significance during the delayed retention testing. Furthermore both studies indicated that where significant differences were found it was the text condition that led to greater retention of information.

That improved learning will result when graphical and text media are combined is a tenet of the CTML, and so it is surprising to find some evidence to the contrary. There are a number of possible reasons for this. As discussed in chapter four, although all of the key points to be assessed by the subsequent questions were included in both the text and the timeline materials, in the text form there was ancillary information involved in the connecting words and phrases that may have proven helpful in subsequent recall. The relatively high number of names, events, dates, and linkages involved in these studies (over 40 in total) compared to those of similar studies may also have been a factor. Korallo (2010b), for example, used only nine items. Some other studies have also found the CTML modality effects to reduce over time, for example Korallo (2010b) found that the learning difference between the virtual reality timeline and the control had disappeared or, in one study, had even reversed

after a delay of 2-6 weeks.

**Hypothesis Four: Individual creation of timelines will result in greater retention of information when compared to provided timelines**

This hypothesis was examined in study four. There is a great deal of research supporting the use of learner constructed materials in learning (Van Meter et al., 2006; Schwamborn et al., 2010; Prangma, 2007) and the generative drawing principle of CTML directly relates this to multimedia examples. The results indicated that the effect of delivery (presented vs created) was significant, but with those who experienced the created materials condition showing a lower average score across the two sessions, than those who were provided with the materials. These results conflict with the hypothesis that suggested self creation would lead to improved scores. Furthermore the results did not show an interaction between the mode and delivery. This suggests that this effect was not limited to the timeline condition. This mixed result highlights the need for further studies, designed with the primary focus of exploring delivery mode. In particular it would be valuable to explore whether the provided materials were too closely matched to the assessment materials, and whether additional guidance during the self creation of resources, such as a teacher would give during a lesson, would be beneficial. The generative drawing principle has also, like many others in CTML been developed in connection with examples from science and engineering where the purpose of a diagram is as an aid to the learner making sense of a process or mechanism. It may have been found here to be inapplicable for the more abstract cases of a timeline or map style of diagram.

**Hypothesis Five: Question response times will have a positive correlation with self reported estimates of cognitive load**

Hypothesis five was explored in studies two and three. This hypothesis was designed to contribute to research examining ways to measure cognitive load (Chen et al., 2016; van Gog & Paas, 2008; van Gog, Kirschner, Kester, &

Paas, 2012), and in particular, to develop measures which do not distract the user, either through the interruptions created by secondary tasks, or by self reporting. The results from study two provided support for this hypothesis, with the self assessed cognitive loads showing a weak to medium correlation for all five types of question, in both the text and timeline conditions. This was also present when considering only the questions which were correctly answered.

### **Hypotheses Summary**

It is clear from the previous discussion that the study results provided mixed support for the five hypotheses. The key findings are that study one provided real world support for hypothesis one indicating that timelines can help students in their classroom assessment. In studies two and three participant response times during comprehension were found to be shorter for the timeline condition than for text, supporting hypothesis two. Investigations into hypothesis three showed significant results going in the opposite direction, however, with text proving beneficial for retention. Similarly, contradictory results were found regarding hypothesis four where it was shown that providing summary materials led to higher scores than when the participants generated their own. Finally hypothesis five was accepted, with reaction times showing some correlation with self reported cognitive load. In the following sections the key research questions which were established in the literature section are reflected upon, discussing the extent to which this thesis helps further the understanding of timelines, learning and cognitive load research.

Hypothesis	Study one	Study two	Study three	Study four
Use of timelines will result in increased comprehension of information when compared to text.	Supported	Not supported	Not supported	
Use of timelines will result in shorter question response times during comprehension of information when compared to text.		Supported	Supported	
Use of timelines will result in greater retention of information when compared to text.		Not supported - Opposite	Not supported - Opposite	Not supported
Individual creation of timelines will result in greater retention of information when compared to provided timelines.				limited, with those creating seeming to do worse, but there was persistence of learning
Question response times will have a positive correlation with self reported estimates of cognitive load.		Supported		

**Table 7.2** – Hypotheses and support from Studies

## **7.5 To what extent can the CTML principles be applied to timelines?**

From reviewing the literature it became apparent that there are a wide range of theories in relation to learning, cognitive load and optimal approaches for using multimedia materials. In this thesis the CTML was chosen as it has an established set of principles which have been repeatedly explored. It was noted that these repeated explorations mostly involved scientific, engineering, and mathematical materials and that there was only limited research exploring timelines and how they may aid in the learning process. As the CTML principles have been shown to be effective for the other subjects, this thesis sought to establish if the same principles could be applied when investigating learning using timelines. In this next section, key principles are considered in turn, along with a discussion of their replicability with the timelines used in these studies.

### **Multimedia Principle: Learning is increased when the materials used contain words and pictures rather than just words**

In general timelines should fit very neatly with the multimedia principle, combining, as they do, graphical elements and associated text in one image, the "Big Picture". They use the placing of events to indicate order, the overall structure and linking lines to show causality, and textual labels to provide event detail. To comprehend and learn from a timeline, therefore, requires the cognitive integration of all of those aspects, in the same manner as other image and text based multimedia. The present studies produced varying results in relation to timelines improving learning, however, with only study one showing a clear benefit from having information provided in a timeline format, while the results from two of the laboratory based studies actually showed greater retention for the text condition. In this research the multimedia principle was not found to work consistently with timelines in the way that many researchers have found for other material types and conditions. Mayer

(2014a) acknowledges that research is beginning to show boundary conditions for some of the CTML principles, and that in some cases, outside of the boundaries, the principle must be reversed. The earliest example of such a reversal, was actually defined as a principle in its own right, the "expertise reversal principle" (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, 2007) in which instructional techniques that are effective for novices become ineffective, and may even harm learning, for more experienced learners. Expertise reversal has been found to apply to a number of different principles; audio/visual modality (Inan et al., 2015), redundancy (Leslie, Low, Jin, & Sweller, 2012), segmentation (Spanjers, Wouters, van Gog, & van Merriënboer, 2011; Blayney et al., 2016), as well as being attributed to overall complexity (Blayney et al., 2016; Kalyuga & Singh, 2016). Boundary conditions less directly related to expertise, that have been found to lead to reversals, include learner control of study time (McCrudden, Hushman, & Marley, 2014), and length of material (Schuler et al., 2012; Schuler, Scheiter, & Gerjets, 2013). Lindow et al. (2011) found the modality effect underlying some of the CTML principles not to be robust. Mayer (2014a) considers that more research is needed regarding boundaries, so it may well be that boundary conditions have been encountered in the studies within this thesis. Some other studies have also found reversals to occur only when longer term assessments have been made, such as those testing retention at a later stage (Schweppe & Rummel, 2012; Schweppe, Eitel, & Rummel, 2015). Whilst this may not explain the lack of support for the principle in immediate retention tests, it may do for the delayed retention in study three.

**Split Attention: Learning is increased when learners do not have to split their attention between several sources of information and Signalling/Cueing: Learning is increased when key information is pointed out by visual cues such as colour coding or bold text**

Timelines should be beneficial to learning as they act as one source with multiple pieces of information, rather than requiring the learner to read through a large amount of text and distill the relevant information (Kalyuga et al., 1999; Florax & Ploetzner, 2010). They can

also be used to highlight pertinent information. These principles were borne out to an extent, as those in the timeline condition spent less time reviewing and understanding the materials, but at the immediate comprehension stage were able to perform at the same level as those who spent longer looking at the text equivalent. This suggests that the timeline may have helped focus their attention, without the need for re-reading and searching for answers. As this did not continue to the retention phases, it is possible to suggest that the timelines help the learner in focusing their attention and comprehending information, but are limited in their ability to support translation into a deeper level of learning. The split attention and signalling/cueing principles were used in the design of materials for the studies (although the split attention principle was knowingly ignored for study four since the study was based around the provision of additional material and so involved two separate items). They were not explored in any comparison study, however.

### **Generative Drawing: Encouraging learners to produce their own diagrams while learning can increase learning**

This principle was explored in the final study, but was not proven. Regardless of whether the participant was asked to write their own notes or produce a timeline, neither group performed better than the groups who were provided with notes or timelines. On reflection it is possible that the provided notes and timelines may have been matched too closely with the assessment materials. However, this is representative of the school education system where revision guides often signpost the key parts of the syllabus that will be tested. In this research at least, self generation of timelines (or text) was not beneficial to this measure of learning. A recent study by Lin et al. (2017), contrasting the learning effects of self generated drawing with repeated reading, and self imagined images, found some support for the principle, but only where prior knowledge was low (which was the case for study four since the material was fictional) but concluded that their sample size (which was similar to that of study four) was too small, and suggested that more study is needed.



## **Timelines and the CTML in summary**

From the previous discussions it appears that the timelines used in the present studies may not fit robustly with the principles of the CTML, since, in some cases they were shown to help the learner, but in others this effect was not sustained. Other researchers have encountered similar variability in results regarding learning from multimedia, when either repeating experiments or extending to different types of material (Korrallo, 2010b; Prangma, 2007), and different aspects of the material (Eitel, Kuhl, Scheiter, & Gerjets, 2014). A number of attempted replications of the modality effect (mostly comparing text with audio) have found similar difficulties (McNeill et al., 2009; Tabbers et al., 2004; Oberfoell & Correia, 2016; Lindow et al., 2011). In a number of cases any effects seen in immediate testing either disappeared or were reversed after delays (Korrallo, 2010b; Prangma, 2007; Schweppe & Rummer, 2016), and studies two and three certainly encountered this. Schweppe and Rummer (2012) also point out that the majority of CTML related studies only use immediate testing in their evaluations and that more work should be performed using delayed testing. In a related manner, when examining effects of both Kulhavy's (1985) Conjoint Retention Hypothesis (CRH) and Paivio's (1991) Dual Coding Theory (DCT), Diana and Webb (1997) found that these effects weakened over time. The studies performed by McNeill et al. (2009) are particularly interesting since, although focused upon comparing modalities of audio, text, and animation rather than text and images used here, they were also using history as the subject. They did not find any modality or redundancy effects. They also highlight an issue with many CTML studies, acknowledged by Mayer (1999), that most of the test material used for research is short. In the studies discussed here, the test material and activities were designed to more closely mimic the amount of work involved in a topic within a school lesson (in the case of study one, that was exactly what was done). The materials all involved at least 30 different items to be remembered, and the study times were all at least 10 minutes (learner controlled in studies two and three, and fixed at 10 minutes for study four). The overall number of items

for participants to learn raises another point of discussion, however. When performing an exploratory study regarding learning from maps (and similar diagrams, some of which can be considered to be close in structure to timelines), Winn and Sutherland (1989) showed that recall performance could be adversely affected by increasing the number of items involved, their maximum number of elements being 20.

It is also important, however, to consider that 'learning' is a complicated term and there are no clear criteria for confirming that learning has occurred. Returning to the literature it is clear that work using the CTML more often focuses on the investigation of "transfer" than on the type of comprehension and retention effects studied here, and it may be that this is an aspect requiring further research using timeline material. Despite the mixed findings, however, the research presented in this thesis provides an initial starting point for future research when looking to extend the CTML into areas outside its common usage in science, engineering, and mathematics.

## **7.6 Should timelines be considered as maps of time or as diagrams?**

The long-standing view from early chronographers (Rosenberg & Grafton, 2013) suggests that timelines are maps of time. The consequential idea that learning from a timeline could engage spatial memory, as in the conjoint retention hypothesis (CRH) (Kulhavy et al., 1985), and seen to apply in some experiments by Korallo (2010b), was considered as a part of the work described here but not directly explored in detail. Much of the work on the CRH relates to geographic maps, where the participants can directly relate the map to their direct experience of the tangible world and some researchers have argued that it is the recognisable features of the map, mimetic icons, and well known symbols, that are stored in visuospatial memory (Nesbit & Adesope, 2006; Griffin & Robinson, 2005; Newbern et al., 1997).

Although the term timeline is seeing increased use (in areas of social media on the Internet

for example) there is still no strong commonality of layout, or symbols, across all timelines, nor much in the way of common icons that would be clearly recognised by all in the case of a timeline. The "edge" features of a map (coasts, and strongly marked internal features, borders, etc.) have been shown to have a stronger effect on memory, with text close to those features being better retained and recalled (Verdi & Kulhavy, 2002; Winn & Sutherland, 1989; Verdi, Stamm, Johnson, & Jamison, 2001). It could be that strong lines in a timeline may engage a similar effect, but this would require further study, perhaps using simpler timelines, varying the direction in connection with the expectations of a mental timeline (Weger & Pratt, 2008; Flumini & Santiago, 2013; Bonato et al., 2012; Bender & Beller, 2014; Aguirre & Santiago, 2017), and exploring the effects of different scales particularly using a non linear scale. Neither the Year 7 (Figure B.1) nor Year 8 timelines (Figure B.2 on page 239) used a linear scale, events were simply placed in order of occurrence. The same approach can be seen in the classroom timeline example in Figure 3.1 on page 64. Exploring different scales (including logarithmic) could provide a useful insight into the relationship between physical timelines, mental timelines and associations with A Theory Of Magnitude (ATOM) (Fabbri et al., 2012; Winter, Marghetis, & Matlock, 2015; Bottini & Casasanto, 2010), particularly with the more logarithmic nature of the cognitive/neural magnitude estimations within ATOM involving the Weber-Fechner law (Nieder & Miller, 2003; Buetti & Walsh, 2009; Bonato et al., 2012).

The timelines in this study have, however, been considered as images in respect of the CTML and it is worth reflecting upon the results that have been found by treating them this way. Study one did find a benefit from using timelines, though the example used, had a richer content (through the inclusion of additional cartoon pictures) than any of the timelines used in the three other studies and it may be that it was this stronger imagery that resulted in the expected CTML effect. It would be interesting to explore the effects of different timeline styles including the incorporation of varying amounts of picture material in a laboratory based experiment.

## **7.7 Can response times be used as an indicator for cognitive load?**

A review of the existing research suggests that learning can be limited by the effects of cognitive load (Sweller & Chandler, 1994; Sweller, 2015; Fiorella & Mayer, 2016; Sweller, 1988, 1994, 2010). The notion of cognitive load, and of the summing of different types of load (intrinsic, extrinsic, germane) (DeLeeuw & Mayer, 2008; Leopold & Mayer, 2015) forms a key foundation for the CTML (Mayer, 2005). This limit is complex, however, with a balancing act needed between providing enough load that the student has to consider the information, and reflect upon it, the desirable difficulty (Lehmann, Goussios, & Seufert, 2016; Maddox & Balota, 2015; Schweppe & Rummer, 2016), and the load being too high and the student being unable to integrate the information into long term storage. A large number of studies have used self report as a measure for cognitive load; however this disrupts the study and impacts the learning process by diverting the participant's attention. Two of the studies reported in this thesis took the opportunity to explore whether response time could be used as a mechanism for judging cognitive load since this could be performed unobtrusively. A comparison with a self report instrument commonly used by other studies suggested that there was a moderate correlation between the two measures and that the response time might, therefore, provide a suitable measure for assessing cognitive load. However the analysis of the response times found a number of confounding effects that would need to be dealt with in order for the approach to be used more generally. The first of these concerned the inherent variability of the tasks being performed. The response times varied depending upon the type of response that the participant needed to make (e.g. keyboard vs mouse), on the number of options to select between, and particularly on the speed-accuracy trade off. More sophisticated models of the interaction between testing and response times such as item response theory (Jabrayilov, Emons, & Sijtsma, 2016; Klein Entink, Kuhn, Hornke, & Fox, 2009) should be explored before deciding upon any

wider adoption of the technique.

## **7.8 Is there a relationship between timeline benefits and spatial abilities, or learning preferences?**

There has been discussion in the literature of spatial ability and learning preferences impacting on how students learn from multimedia (Massa & Mayer, 2006; Bednarz & Lee, 2011; Pazzaglia & Moe, 2013; Brunken et al., 2003; Hoffler, 2010). Studies two and three required distractor tasks interposed between the main parts of the experiment, and these were arranged to include measures relating to spatial ability and learner preference. (This would enable additional analysis to be performed to determine whether performance in the two conditions was mediated by the participant's spatial ability or self reported multimedia learning preferences). In neither study was there a relationship found between the participant's study performance and their spatial ability, or learner preferences. This could be because the assessment instruments chosen were not as appropriate as originally believed, and different instruments may have aligned better with timelines as the multimedia material, or it could be that the ability range of participants in the studies where this was assessed was narrow, and thus the effects were not seen. Hoffler (2010) also concluded that the studies where differences had been seen mostly concerned the effects of animation, which was not involved in the studies described here. Despite this, the research described in this thesis does add to the body of work on spatial ability and learning, and may prove useful to future researchers.

## **7.9 Reflections, implications, and future work**

During earlier discussions around multimedia studies and associated literature, the observation was made that most of the existing work related to science and technology. It

must also be acknowledged that the studies within this thesis also only examined a small fraction of possible history related examples. Five different styles of timeline were used but there are very many other styles that could have been explored. The measures of learning were limited to comprehension and retention and the assessments in studies two to four only used multiple choice testing in order to be completely objective, whereas some similar studies have used essays or other forms of marked work, together with suitable inter rater reliability assessments to mitigate subjectivity. These limitations are addressed in later suggestions for future work.

As discussed in the introduction and literature chapters, the most immediate real world use of learning from timelines concerns their use in schools, in particular for teaching history. In this regard the school based study did show a benefit from the use of timelines for one year group. Studies two and three each had a scenario where text had an advantage, but it might be considered not to be a very robust result since the effect moved between sessions for the two studies. Where both studies did agree, however, was that for similar comprehension results the overall study times were lower for the timeline condition, so perhaps timelines may be more readily searched for information. This has relevance for school learning where the analysis of three years of teaching materials (performed by the researcher as a precursor to study one), the associated discussions with teachers, and researcher observation of a history class after the study, showed just what a tremendous pace of learning has to be kept up to cover the expected syllabus. Anything that speeds up comprehension of the material to be learned would be a benefit.

In industry, timeline-like representations are often found in project planning (e.g. Gantt charts (Burkhard, Meier, Rodgers, Smis, & Stott, 2005; Geraldi & Lechter, 2012)) and, although many people are used to reading these, they can be hard to comprehend (Zimoch, Pryss, Probst, Schlee, & Reichert, 2017; Burkhard et al., 2005), and so research that improves understanding of the best approach to presentation of timelines may be of benefit. It is perhaps in everyday life, though, that research into the comprehension of temporal data

through timelines could have its widest benefit. Only a few years ago many people would be unfamiliar with the term timeline, yet now, with its frequent usage in connection with social media on the Internet, it is likely to be much more common. Social media timelines are not the same as the learning timelines discussed here but, as they develop to help users cope with the rapid growth in the amount of data (often time related) that people have to deal with, they may well produce innovative approaches that could also be used in learning. Such approaches could be a useful area for further research.

A number of other areas for future research have been identified during the discussion of the studies within this thesis. Starting from the viewpoint of educational use of timelines, all of the studies have used relatively complex timelines. These timelines involving mixtures of implied and/or diagrammatic lines, associated text, and, in some cases, images. The complexity has meant that it has not been possible to dissociate the overall imagery effects under the CTML (Mayer, 2014a; Schuler, Arndt, & Scheiter, 2015; Schuler, Scheiter, & Gerjets, 2010; Eitel, Scheiter, & Schuler, 2013; Richter et al., 2016; Merkt & Schwan, 2017; Fenesi & Kim, 2014), from any benefits that may have resulted from connection between a timeline and mental timelines (Bonato et al., 2012; Boroditsky et al., 2011; Hartmann & Mast, 2012; Arzy et al., 2009; Droit-Volet & Coull, 2015). A study using very simple timelines, perhaps exploring different directions of increasing magnitude could provide a useful foundation for other studies.

Drawing timelines with the oldest event on the left is a common approach in modern chronographics (Davis, 2012), and also considered to be the direction of mental timelines, at least in western cultures with a left-right reading direction (Pitt & Casasanto, 2016; Nunez & Cooperrider, 2013; Vallesi et al., 2011; de la Vega, Eikmeier, Ulrich, & Kaup, 2016; Weger & Pratt, 2008; Vallesi et al., 2008), however cultural differences have also been found (Yang & Sun, 2016; Bergen & Chan Lau, 2012; Fuhrman & Boroditsky, 2007; Majid, Gaby, & Boroditsky, 2014; Magnani & Musetti, 2017). Different subjects such as genealogy and geological timelines sometimes use a vertical time axis (Mitchell, 2014), and studies

using virtual reality have used a primarily forwards and backwards direction (Foreman et al., 2008; Korallo, 2010a). This could then be extended to studying (again using very simple timelines) cases where, as in the Year 9 example from Study One in Figure B.3 on page 240, two dimensions are used simultaneously. The more common approach of including multiple lines in the same diagram, as seen in the materials for the other studies, could then be investigated, followed by examination of the effects of varying colour and other highlighting, and mimetic icons.

In some timelines the key aspect being conveyed concerns the sequence of events, with no significance placed upon the spacing between the events. There is either no scale at all, or a non-linear scale is used. All three timelines in study one used this approach (for all aspects of the timeline in Y7 and Y8 and for the vertical dimension in Y9). A study, again based upon simple timelines, to examine how gaps and scales are perceived and remembered could investigate any links with A Theory Of Magnitude, especially concerning any scaling effects of the Weber-Fechner law (Nieder & Miller, 2003; Buetti & Walsh, 2009; Bonato et al., 2012). Geological timelines for example often need to resort to non-linear scaling because they have to cover very large periods and some studies have shown student perceptions of the numbers involved and the extension to timeline representations, to be non-linear (Dehaene, Izard, Spelke, & Pica, 2008; Libarkin, Kurdziel, & Anderson, 2007). Two of the timelines used in study one had non linear scales but, again, this was incidental to the study and not able to be investigated.

The use of eye tracking when performing research into multimedia learning is also becoming more prevalent (Lin et al., 2017; Mayer, 2010; Ponce & Mayer, 2014) as the devices used become more readily available and less cumbersome. A study of timeline usage incorporating eye tracking as a measure would form a useful reference point for improving their design as multimedia items and, since it may well be possible to measure pupil dilation and blink rate at the same time, it may be possible to produce a study that also incorporates this as an intrinsic measure of cognitive load (Zekveld et al., 2014; Chen &



Epps, 2013) as the learning takes place.

## **7.10 Final comments**

While the results of these studies have been varied and, as ever, future research would be beneficial, this thesis has explored a wide range of aspects linked to timelines and learning, and does provide a strong foundation for future research to help fill in the existing research gaps. As technology, and education, are both ever changing areas, refining the multimedia principles to keep up with developments and improving the understanding of where the boundaries of applicability lie will be an important task. A final suggestion for future work, links back to the work of Korallo (2010b), which was inspirational from an early stage of the work on this thesis. With the many improvements and innovations in technology that have taken place, it could be that the experience of walking through time in fully immersive virtual reality becomes the best way to understand history!

## References

- Abdul-Rahman, S.-S. & du Boulay, B. (2014). Learning programming via worked-examples: Relation of learning styles to cognitive load. *Computers in Human Behavior*, 30, 286–298. doi:10.1016/j.chb.2013.09.007
- Abel, R. R. & Kulhavy, R. W. (1989). Associating map features and related prose in memory. *Contemporary Educational Psychology*, 14(1), 33–48. doi:10.1016/0361-476X(89)90004-0
- Adesope, O. O. & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology*, 104(1), 250–263. doi:10.1037/a0026147
- Aguirre, R. & Santiago, J. (2017). Do potential past and future events activate the Left-Right Mental Timeline? *Psicologica*, 38(2).
- Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). *Visualization of Time-Oriented Data*. Human-Computer Interaction Series. London: Springer London.
- Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33(2), 131–152. doi:10.1016/S0360-1315(99)00029-9
- Ainsworth, S., Galpin, J., & Musgrove, S. (2007). Learning about dynamic systems by drawing for yourself and for others. In *EARLI conference, Budapest*. European Association for Research on Learning and Instruction.
- Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, 333(6046), 1096–1097.

- Ainsworth, S. & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction, 14*(3), 241–255. doi:10.1016/j.learninstruc.2004.06.002
- Alards-Tomalin, D., Leboe-McGowan, J. P., Shaw, J. D. M., & Leboe-McGowan, L. C. (2014). The effects of numerical magnitude, size, and color saturation on perceived interval duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 555–566. doi:10.1037/a0035031
- Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How Subject-Matter Knowledge Affects Recall and Interest. *American Educational Research Journal, 31*(2), 313. doi:10.2307/1163312
- Alleman, J. & Brophy, J. (2003). History Is Alive: Teaching Young Children about Changes over Time. *The Social Studies, 94*(3), 107–110. doi:10.1080/00377990309600191
- Allen, G. L., Miller Cowan, C. R., & Power, H. (2006). Acquiring information from simple weather maps: Influences of domain-specific knowledge and general visual-spatial abilities. *Learning and Individual Differences, 16*(4), 337–349. doi:10.1016/j.lindif.2007.01.003
- Amorim, M.-A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations: "Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General, 135*(3), 327–347. doi:10.1037/0096-3445.135.3.327
- An, D. & Carr, M. (2017). Learning styles theory fails to explain learning and achievement: Recommendations for alternative approaches. *Personality and Individual Differences, 116*, 410–416. doi:10.1016/j.paid.2017.04.050
- Anderman, E. M., Gimbert, B., O'Connell, A. A., & Riegel, L. (2015). Approaches to academic growth assessment. *British Journal of Educational Psychology, 85*(2), 138–153. doi:10.1111/bjep.12053
- Anderson, J. R. (1981). Interference: The relationship between response latency and response accuracy. *Journal of Experimental Psychology: Human Learning and Memory, 7*(5), 326–343. doi:10.1037/0278-7393.7.5.326

- Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2(6), 396–408. doi:10.1111/j.1467-9280.1991.tb00174.x
- Andrade, J., May, J., Deeptose, C., Baugh, S.-J., & Ganis, G. (2013). Assessing Vividness of Mental Imagery: The Plymouth Sensory Imagery Questionnaire. *British Journal of Psychology*, 105. doi:10.1111/bjop.12050
- Andre, P., Wilson, M. L., Russell, A., Smith, D. A., & Owens, A. (2007). Continuum: designing timelines for hierarchies, relationships and scale. In *Proceedings of the 20th annual ACM symposium on User interface software and technology* (pp. 101–110). ACM. doi:10.1145/1294211.1294229
- Anvari, F., Tran, H. M. T., & Kavakli, M. (2013). Using Cognitive Load Measurement and Spatial Ability Test to Identify Talented Students in Three-Dimensional Computer Graphics Programming. *International Journal of Information and Education Technology*, 94–99. doi:10.7763/IJiet.2013.V3.241
- Ari, F., Flores, R., Inan, F. A., Cheon, J., Crooks, S. M., Paniukov, D., & Kurucay, M. (2014). The effects of verbally redundant information on student learning: An instance of reverse redundancy. *Computers & Education*, 76, 199–204. doi:10.1016/j.compedu.2014.04.002
- Arzy, S., Adi-Japha, E., & Blanke, O. (2009). The mental time line: An analogue of the mental number line in the mapping of life events. *Consciousness and Cognition*, 18(3), 781–785. doi:10.1016/j.concog.2009.05.007
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The Psychology of learning and motivation: Advances in research and theory* (vol. 2). 89–105. doi:10.1016/s0079-7421(08)60422-3
- Aue, W. R., Criss, A. H., & Fischetti, N. W. (2012). Associative information in memory: Evidence from cued recall. *Journal of Memory and Language*, 66(1), 109–122. doi:10.1016/j.jml.2011.08.002

- Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, 16(5), 389–400.  
doi:10.1016/j.learninstruc.2006.09.001
- Ayres, P. (2015). State-of-the-Art Research into Multimedia Learning: A Commentary on Mayer's Handbook of Multimedia Learning: Multimedia learning. *Applied Cognitive Psychology*, 29(4), 631–636. doi:10.1002/acp.3142
- Baddeley, A. D. & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, 129(1), 126–145.  
doi:10.1037//0096-3445.129.1.126
- Baddeley, A. D. & Hitch, G. J. (1975). Working Memory. *Psychology of Learning and Motivation*, 8, 47–87. doi:10.1016/S0079-7421(08)60452-1
- Baddeley, A. D. & Hitch, G. J. (2000). Development of Working Memory: Should the Pascual-Leone and the Baddeley and Hitch Models Be Merged? *Journal of Experimental Child Psychology*, 77(2), 128–137. doi:10.1006/jecp.2000.2592
- Baddeley, A. D., Kopelman, M. D., & Wilson, B. A. (2004). *The Essential Handbook of Memory Disorders for Clinicians*. Chichester, England; Hoboken, NJ, USA: J. Wiley.
- Baguley, T. (2012). *Serious Stats: A guide to advanced statistics for the behavioral sciences*. Macmillan Education UK.
- Barnett, S. M. & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637.  
doi:10.1037//0033-2909.128.4.612
- Bartlett, F. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press. Cambridge University Press.
- Bauhoff, V., Huff, M., & Schwan, S. (2012). Distance Matters: Spatial Contiguity Effects as Trade-Off between Gaze Switches and Memory Load: Spatial contiguity. *Applied Cognitive Psychology*, 26(6), 863–871. doi:10.1002/acp.2887

- Bausenhardt, K. M. (2014). Mechanisms of Multisensory Integration in the Time Domain. *Procedia - Social and Behavioral Sciences*, 126, 4–5.  
doi:10.1016/j.sbspro.2014.02.296
- Bednarz, R. S. & Lee, J. (2011). The components of spatial thinking: empirical evidence. *Procedia - Social and Behavioral Sciences*, 21, 103–107.  
doi:10.1016/j.sbspro.2011.07.048
- Beilin, H. (1992). Piaget's enduring contribution to developmental psychology. *Developmental Psychology*, 28(2), 191. doi:10.1037/0012-1649.28.2.191
- Bender, A. & Beller, S. (2014). Mapping spatial frames of reference onto time: A review of theoretical accounts and empirical findings. *Cognition*, 132(3), 342–382.  
doi:10.1016/j.cognition.2014.03.016
- Bender, A., Rothe-Wulf, A., Huther, L., & Beller, S. (2012). Moving Forward in Space and Time: How Strong is the Conceptual Link between Spatial and Temporal Frames of Reference? *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00486
- Bennett, P., Fraser, M., & Balaam, M. (2012). ChronoTape: tangible timelines for family history. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (pp. 49–56). ACM. doi:10.1145/2148131.2148144
- Bergen, B. K. & Chan Lau, T. T. (2012). Writing Direction Affects How People Map Space Onto Time. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00109
- Besson, G., Ceccaldi, M., Didic, M., & Barbeau, E. J. (2012). The speed of visual recognition memory. *Visual Cognition*, 20(10), 1131–1152.  
doi:10.1080/13506285.2012.724034
- Beste, C., Heil, M., & Konrad, C. (2010). Individual differences in ERPs during mental rotation of characters: Lateralization, and performance level. *Brain and Cognition*, 72(2), 238–243. doi:10.1016/j.bandc.2009.09.005

- Bizo, L. A., Chu, J. Y., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber's law in time perception and production. *Behavioural Processes*, 71(2-3), 201–210.  
doi:10.1016/j.beproc.2005.11.006
- Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.  
doi:10.1080/0969595980050102
- Blatchford, P., Russell, A., Bassett, P., Brown, P., & Martin, C. (2007). The role and effects of teaching assistants in English primary schools (Years 4 to 6) 2000-2003. Results from the Class Size and Pupil-Adult Ratios (CSPAR) KS2 Project. *British Educational Research Journal*, 33(1), 5–26. doi:10.1080/01411920601104292
- Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated-interactive elements effect and levels of learner expertise: Experimental evidence from an accountancy class. *Instructional Science*, 38(3), 277–287.  
doi:10.1007/s11251-009-9105-x
- Blayney, P., Kalyuga, S., & Sweller, J. (2016). The impact of complexity on the expertise reversal effect: experimental evidence from testing accounting students. *Educational Psychology*, 36(10), 1868–1885. doi:10.1080/01443410.2015.1051949
- Bobek, E. & Tversky, B. (2016). Creating visual explanations improves learning. *Cognitive Research: Principles and Implications*, 1(1). doi:10.1186/s41235-016-0031-6
- Bolls, P. D. (2002). I Can Hear You, but Can I See You?: The Use of Visual Cognition During Exposure to High-Imagery Radio Advertisements. *Communication Research*, 29(5), 537–563. doi:10.1177/009365002236194
- Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience & Biobehavioral Reviews*, 36(10), 2257–2273.  
doi:10.1016/j.neubiorev.2012.08.007
- Borella, E., Ghisletta, P., & de Ribaupierre, A. (2011). Age Differences in Text Processing: The Role of Working Memory, Inhibition, and Processing Speed. *The Journals of*

- Gerontology Series B: Psychological Sciences and Social Sciences*, 66B(3), 311–320.  
doi:10.1093/geronb/gbr002
- Borella, E., Meneghetti, C., Ronconi, L., & De Beni, R. (2014). Spatial abilities across the adult life span. *Developmental Psychology*, 50(2), 384–392. doi:10.1037/a0033818
- Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, 75(1), 1–28. doi:10.1016/S0010-0277(99)00073-6
- Boroditsky, L. (2011). How Languages Construct Time. In *Space, Time and Number in the Brain* (pp. 333–341). London: Elsevier Academic Press.
- Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, 118(1), 123–129.  
doi:10.1016/j.cognition.2010.09.010
- Boroditsky, L. & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, 13(2), 185–189. doi:10.1111/1467-9280.00434
- Borra, C., Iacovou, M., & Sevilla, A. (2012). The effect of breastfeeding on children's cognitive and noncognitive development. *Labour Economics*, 19(4), 496–515.  
doi:10.1016/j.labeco.2012.05.007
- Borst, G. (2013). Fear improves mental rotation of low-spatial-frequency visual representation. *Emotion*, 13(5), 811–816. doi:10.1037/a0033625
- Bottini, R. & Casasanto, D. (2010). Implicit spatial length modulates time estimates, but not vice versa. In *International Conference on Spatial Cognition* (pp. 152–162). Springer.  
doi:10.1007/978-3-642-14749-4\_15
- Bottini, R. & Casasanto, D. (2013). Space and time in the child's mind: metaphoric or ATOMIC? *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00803
- Brockmole, J. R. & Logie, R. H. (2013). Age-Related Change in Visual Working Memory: A Study of 55,753 Participants Aged 8-75. *Frontiers in Psychology*, 4.  
doi:10.3389/fpsyg.2013.00012



- Brophy, J. & Alleman, J. (2003). History Is Alive: Teaching Young Children about Changes over Time. *Social Studies*, 94(3), 107–110. doi:10.1080/00377990309600191
- Brosnan, M., Walker, I., & Collomosse, J. (2010). The effect of explicitly varying the proportion of "same" and "different" responses on sex differences in the Shepard and Metzler mental rotation task. *European Journal of Cognitive Psychology*, 22(2), 172–189. doi:10.1080/09541440902743710
- Bruner, J. (2006). *In Search of Pedagogy Volume I: The Selected Works of Jerome Bruner, 1957-1978*. Taylor & Francis.
- Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 53–61. doi:10.1207/S15326985EP3801\_7
- Brunken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In *Cognitive Load Theory*. UK: Cambridge University Press.
- Bueti, D. & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1525), 1831–1840. doi:10.1098/rstb.2009.0028
- Burkhard, R. A., Meier, M., Rodgers, P., Smis, M. T., & Stott, J. (2005). Knowledge visualization: A comparative study between project tube maps and gantt charts. *Proceedings of I-Know'05*, 388–395.
- Burny, E., Valcke, M., & Desoete, A. (2009). Towards an Agenda for Studying Learning and Instruction Focusing on Time-Related Competences in Children. *Educational Studies*, 35(5), 481–492. doi:10.1080/03055690902879093
- Butcher, K. R. (2014). The multimedia principle. In *The Cambridge handbook of multimedia learning* (pp. 174–205). UK: Cambridge University Press.
- Butler, M. H., Gilbert, J., Seaborne, A., & Smathers, K. (2004). Data conversion, extraction and record linkage using XML and RDF tools in Project SIMILE. *Technical Report HPL-2004-147 Bristol, UK, HP Labs*.

- CACE. (1968). *Children and Their Primary Schools: The report*. Children and Their Primary Schools: A Report of the Central Advisory Council for Education. H.M.S.O.
- Canham, M. & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, 20(2), 155–166. doi:10.1016/j.learninstruc.2009.02.014
- Carney, R. N. & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14(1), 5–26. doi:10.1023/A:1013176309260
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, 23(6), 760–771. doi:10.1002/acp.1507
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Casasanto, D. & Boroditsky, L. (2003). Do we think about time in terms of space. In *Proceedings of the 25th annual meeting of the Cognitive Science Society* (pp. 216–221). doi:10.1207/s15516709cog2702\_11
- Casasanto, D. & Boroditsky, L. (2008). Time in the Mind: Using Space to Think About Time. *Cognition*, 106(2), 579–593. doi:10.1016/j.cognition.2007.03.004
- Casasanto, D. & Bottini, R. (2010). Can Mirror-Reading Reverse the Flow of Time? In C. Holscher, T. F. Shipley, M. Olivetti Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), *Spatial Cognition VII* (pp. 335–345). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Cepeda, N. J., Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing Effects in Learning A Temporal Ridgeline of Optimal Retention. *Psychological Science*. doi:10.1111/j.1467-9280.2008.02209.x
- Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, S. Miller, & V. Ageyev (Eds.), *Vygotsky's*

- educational theory in cultural context* (pp. 39–64). Cambridge: Cambridge University Press.
- Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, 18(1), 49–57. doi:10.1080/09658210903405737
- Chen, F., Ruiz, N., Choi, E., Epps, J., Khawaja, M. A., Taib, R., . . . Wang, Y. (2012). Multimodal behavior and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems*, 2(4), 1–36. doi:10.1145/2395123.2395127
- Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust Multimodal Cognitive Load Measurement*. Springer.
- Chen, S. & Epps, J. (2013). Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer Methods and Programs in Biomedicine*, 110(2), 111–124. doi:10.1016/j.cmpb.2012.10.021
- Chen, S.-H., Huang, M.-H., & Chen, D.-Z. (2012). Identifying and visualizing technology evolution: A case study of smart grid technology. *Technological Forecasting and Social Change*, 79(6), 1099–1110. doi:10.1016/j.techfore.2011.12.011
- Cheng, P. C.-H., Lowe, R. K., & Scaife, M. (2001). Cognitive science approaches to understanding diagrammatic representations. *Artificial Intelligence Review*, 15, 79–94.
- Christian, D. & McNeill, W. (2011). *Maps of Time: An Introduction to Big History*. California World History Library. University of California Press.
- Cinan, S., Atalay, D., Sisman, S., Basbug, G., Dervent-Ozbek, S., Teoman, D. D., . . . Yurtsever, O. D. (2007). Memory for object locations: Priority effect and sex differences in associative spatial learning. *Learning and Motivation*, 38(4), 326–341. doi:10.1016/j.lmot.2007.01.002
- Clare, J. D. (2002). Tudor Religion. Retrieved July 25, 2017, from [http://www.johndclare.net/KS3/TudorReligion/Tudor\\_religion.htm](http://www.johndclare.net/KS3/TudorReligion/Tudor_religion.htm)

- Clark, H. H. (1973). Space, time, semantics, and the child. In *Cognitive development and the acquisition of language* (pp. 27–63). New York: Academic Press.
- Clark, J. & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210. doi:10.1007/bf01320076
- Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning*. London: Learning and Skills Research Centre.
- Colom, R., Contreras, M. J., Botella, J., & Santacreu, J. (2002). Vehicles of spatial ability. *Personality and Individual Differences*, 32(5), 903–912.  
doi:10.1016/S0191-8869(01)00095-2
- Coluccia, E. & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, 24(3), 329–340.  
doi:10.1016/j.jenvp.2004.08.006
- Cooper, H. (1994). Historical thinking and cognitive development in the teaching of history. In H. Bourdillon (Ed.), *Teaching History* (pp. 101–121). Routledge.
- Cooper, L. A. & Shepard, R. N. (1973). The time required to prepare for a rotated stimulus. *Memory & Cognition*, 1(3), 246–250. doi:10.3758/BF03198104
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104(2), 163. doi:10.1037/0033-2909.104.2.163
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(01), 87–185.  
doi:10.1017/s0140525x01003922
- Cowan, N. (2014). Working Memory Underpins Cognitive Development, Learning, and Education. *Educational Psychology Review*, 26(2), 197–223.  
doi:10.1007/s10648-013-9246-y
- Craik, F. I. (2002). Levels of processing: Past, present... and future? *Memory*, 10(5-6), 305–318. doi:10.1080/09658210244000135

- Craik, F. I. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.  
doi:10.1016/S0022-5371(72)80001-X
- Cranford, K. N., Tiettmeyer, J. M., Chuprinko, B. C., Jordan, S., & Grove, N. P. (2014). Measuring Load on Working Memory: The Use of Heart Rate as a Means of Measuring Chemistry Students' Cognitive Load. *Journal of Chemical Education*, 91(5), 641–647. doi:10.1021/ed400576n
- Croke, B. (1982). The Originality of Eusebius Chronicle. *The American Journal of Philology*, 103(2), 195. doi:10.2307/294250
- Cruse, I. (2011). Debate on 20 October: teaching of history in schools. *House of Lords Library Note UK*.
- Crutcher, R. J. & Beer, J. M. (2011). An auditory analog of the picture superiority effect. *Memory & Cognition*, 39(1), 63–74. doi:10.3758/s13421-010-0015-6
- Cuevas, J. (2015). Is learning styles-based instruction effective? A comprehensive analysis of recent research on learning styles. *Theory and Research in Education*, 13(3), 308–333. doi:10.1177/1477878515606621
- D'Agostino, P. R., O'Neill, B. J., & Paivio, A. (1977). Memory for pictures and words as a function of level of processing: Depth or dual coding? *Memory & Cognition*, 5(2), 252–256. doi:10.3758/BF03197370
- Davis, S. B. (2012). History on the line: time as dimension. *Design Issues*, 28(4), 4–17.
- Davis, S. B., Bevan, E., & Kudikov, A. (2013). Just in time: defining historical chronographics. In *Electronic Visualisation in Arts and Culture* (pp. 243–257). Springer.
- Dawson, I. (2004). Time for chronology. *Teaching History, Historical Association*, 117, 14–24.
- Dawson, I. (2011). Seeing the bigger picture - developing chronological understanding. In *London History Forum Summary*. London.

- de la Vega, I., Eikmeier, V., Ulrich, R., & Kaup, B. (2016). The Mental Timeline in a Crossed-Hands Paradigm: A Matter of Instruction. *Experimental Psychology*, 63(6), 326–332. doi:10.1027/1618-3169/a000342
- Dean, G. M. & Morris, P. E. (2003). The relationship between self-reports of imagery and spatial ability. *British Journal of Psychology*, 94(2), 245–273. doi:10.1348/000712603321661912
- Dean, R. S. & Kulhavy, R. W. (1981). Influence of spatial organization in prose learning. *Journal of Educational Psychology*, 73(1), 57. doi:10.1037//0022-0663.73.1.57
- Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, 71(1), 1–27. doi:10.3102/00346543071001001
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371. doi:10.1037/0096-3445.122.3.371
- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *science*, 320(5880), 1217–1220.
- DeLeeuw, K. E. & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, 100(1), 223–234. doi:10.1037/0022-0663.100.1.223
- Deng, Z., Chen, Y., Zhu, X., & Li, Y. (2016). The effect of working memory load on the SNARC effect: Maybe tasks have a word to say. *Memory & Cognition*. doi:10.3758/s13421-016-0676-x
- Department for Education. (2013a). History programmes of study Key Stage 3 National Curriculum in England. UK Government.
- Department for Education. (2013b). History Programmes of Study Key Stages 1 and 2 National Curriculum in England. UK Government.

- DeVries, R. (2000). Vygotsky, Piaget, and education A reciprocal assimilation of theories and educational practices. *New Ideas in Psychology*, 18, 187–213.  
doi:10.1016/s0732-118x(00)00008-8
- DfE. (2011). *Education\_2011\_how do pupils progress during Key Stages 2 and 3* (Research No. DFE-RR096). Department for Education UK Government.
- Diana, E. M. & Webb, J. M. (1997). Using geographic maps in classrooms: The conjoint influence of individual differences and dual coding on learning facts. *Learning and Individual Differences*, 9(3), 195–214. doi:10.1016/S1041-6080(97)90006-6
- Dijkstra, K. & Moerman, E. M. (2012). Effects of modality on memory for original and misleading information. *Acta Psychologica*, 140(1), 58–63.  
doi:10.1016/j.actpsy.2012.02.003
- DiSalle, R. (2006). *Understanding Space-Time: The Philosophical Development of Physics from Newton to Einstein*. Cambridge University Press.
- Doherty, P. E. (2016). A study of the effects of spatial contiguity and hierarchically structured headings in a shipboard operating and maintenance manual. *WMU Journal of Maritime Affairs*, 15(1), 97–125. doi:10.1007/s13437-015-0090-4
- D'Oliveira, T. C. (2004). Dynamic Spatial Ability: An Exploratory Analysis and a Confirmatory Study. *The International Journal of Aviation Psychology*, 14(1), 19–38.  
doi:10.1207/s15327108ijap1401\_2
- Doyle, R. A. & Voyer, D. (2013). Bodies and occlusion: Item types, cognitive processes, and gender differences in mental rotation. *The Quarterly Journal of Experimental Psychology*, 66(4), 801–815. doi:10.1080/17470218.2012.719529
- Droit-Volet, S. & Coull, J. (2015). The developmental emergence of the mental time-line: spatial and numerical distortion of time judgement. *PLOS ONE*, 10(7), e0130465.
- Duffy, S. E. (2014). The Role of Cultural Artifacts in the Interpretation of Metaphorical Expressions About Time. *Metaphor and Symbol*, 29(2), 94–112.  
doi:10.1080/10926488.2014.889989

- Ebbinghaus, H. (1885). *Memory; a Contribution to Experimental Psychology*. Martino Publishing.
- Eilam, B. & Ben-Peretz, M. (2010). Revisiting curriculum inquiry: the role of visual representations. *Journal of Curriculum Studies*, 42(6), 751–774.  
doi:10.1080/00220272.2010.497192
- Eilam, B. & Poyas, Y. (2008). Learning with multiple representations: Extending multimedia learning beyond the lab. *Learning and Instruction*, 18(4), 368–378.  
doi:10.1016/j.learninstruc.2007.07.003
- Einstein, G. O., McDaniel, M. A., Owen, P. D., & Cote, N. C. (1990). Encoding and recall of texts: The importance of material appropriate processing. *Journal of Memory and Language*, 29(5), 566–581.
- Eitel, A., Kuhl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency Meets Cognitive Load in Multimedia Learning: Does Harder-to-Read Mean Better-to-Understand?: Disfluency in multimedia learning. *Applied Cognitive Psychology*, 28(4), 488–501.  
doi:10.1002/acp.3004
- Eitel, A., Scheiter, K., & Schuler, A. (2013). How Inspecting a Picture Affects Processing of Text in Multimedia Learning: How picture inspection fosters text comprehension. *Applied Cognitive Psychology*, 27(4), 451–461. doi:10.1002/acp.2922
- Ekuni, R., Vaz, L. J., & Bueno, O. F. A. (2011). Levels of processing: The evolution of a framework. *Psychology & Neuroscience*, 4(3), 333–339.  
doi:10.3922/j.psns.2011.3.006
- Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A Large-Scale Model of the Functioning Brain. *Science*, 338(6111), 1202–1205. doi:10.1126/science.1225266
- Ellis, P. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press.



- Erceg-Hurn, D. M. & Miroseovich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, 63(7), 591–601. doi:10.1037/0003-066X.63.7.591
- Evans, V., Bergen, B., & Zinken, J. (2007). *The Cognitive Linguistics Reader*. Advances in cognitive linguistics. Equinox.
- Fabbri, M., Cancellieri, J., & Natale, V. (2012). The A Theory Of Magnitude (ATOM) model in temporal perception and reproduction tasks. *Acta Psychologica*, 139(1), 111–123. doi:10.1016/j.actpsy.2011.09.006
- Fabbri, M., Cellini, N., Martoni, M., Tonetti, L., & Natale, V. (2013). Perceptual and motor congruency effects in time-space association. *Attention, Perception, & Psychophysics*, 75(8), 1840–1851. doi:10.3758/s13414-013-0519-9
- Fagerland, M. W. & Sandvik, L. (2009). The Wilcoxon-Mann-Whitney test under scrutiny. *Statistics in Medicine*, 28(10), 1487–1497. doi:10.1002/sim.3561
- Farr, J. N., Jenkins, J. J., & Paterson, D. G. (1951). Simplification of Flesch Reading Ease Formula. *Journal of Applied Psychology*, 35(5), 333–337. doi:10.1037/h0062427
- Farrell, P., Dyson, A., Polat, F., Hutcheson, G., & Gallannaugh, F. (2007). The relationship between inclusion and academic achievement in English mainstream schools. *School Effectiveness and School Improvement*, 18(3), 335–352. doi:10.1080/09243450701442746
- Fenesi, B. & Kim, J. A. (2014). Learners misperceive the benefits of redundant text in multimedia learning. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.00710
- Ferguson, R. P. & Martin, P. (1983). Long-term temporal estimation in humans. *Perception & psychophysics*, 33(6), 585–592.
- Ferstl, E. C., Israel, L., & Putzar, L. (2017). Humor Facilitates Text Comprehension: Evidence from Eye Movements. *Discourse Processes*, 54(4), 259–284. doi:10.1080/0163853X.2015.1131583
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications.

- Fiorella, L. & Mayer, R. E. (2016). Eight Ways to Promote Generative Learning. *Educational Psychology Review*, 28(4), 717–741. doi:10.1007/s10648-015-9348-9
- Fischer, M. H. & Shaki, S. (2014). Spatial associations in numerical cognition—From single digits to arithmetic. *The Quarterly Journal of Experimental Psychology*, 67(8), 1461–1483. doi:10.1080/17470218.2014.927515
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221. doi:10.1037/h0057532
- Florax, M. & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction*, 20(3), 216–224. doi:10.1016/j.learninstruc.2009.02.021
- Flumini, A. & Santiago, J. (2013). Time (also) flies from left to right... if it is needed. *Proceedings of the 35th annual meeting of the Cognitive Science Society*, 2315–2320.
- Foreman, N., Boyd-Davis, S., Moar, M., Korallo, L., & Chappell, E. (2008). Can virtual environments enhance the learning of historical chronology? *Instructional Science*, 36(2), 155–173. doi:10.1007/s11251-007-9024-7
- Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language*, 62(1), 1–18. doi:10.1016/j.jml.2009.09.004
- Frey, B. & Sutton, J. M. (2010). A model for developing multimedia learning projects. *Journal of Online Learning and Teaching*, 6(2), 491.
- Friedman, W. (1982). The Nature and Development of Time Concepts in Children. In *The Developmental psychology of time*. Developmental Psychology series. Academic Press.
- Fuhrman, O. & Boroditsky, L. (2007). Mental time-lines follow writing direction: Comparing English and Hebrew speakers. In *Proceedings of the Cognitive Science Society* (Vol. 29).

- Fulda, J., Brehmel, M., & Munzner, T. (2016). TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 300–309. doi:10.1109/TVCG.2015.2467531
- Galton, F. (1880). *Visualised Numerals*. Harrison and Sons, printers.
- Ganis, G. & Kievit, R. (2015). A New Set of Three-Dimensional Shapes for Investigating Mental Rotation Processes: Validation Data and Stimulus Set. *Journal of Open Psychology Data*, 3. doi:http://dx.doi.org/10.5334/jopd.ai
- Gardony, A. L., Taylor, H. A., & Brunye, T. T. (2014). What Does Physical Rotation Reveal About Mental Rotation? *Psychological Science*, 25(2), 605–612. doi:10.1177/0956797613503174
- Gemmell, J., Bell, G., Lueder, R., Drucker, S., & Wong, C. (2002). MyLifeBits: fulfilling the Memex vision. In *Proceedings of the tenth ACM international conference on Multimedia* (pp. 235–238). ACM. doi:10.1145/641043.641053
- Geraldi, J. & Lechter, T. (2012). Gantt charts revisited: A critical analysis of its roots and implications to the management of projects today. *International Journal of Managing Projects in Business*, 5(4), 578–594. doi:10.1108/17538371211268889
- Gibson, J. J. (1975). Events are perceivable but time is not. In *The study of time II* (pp. 295–301). Springer.
- Gil, L., Braten, I., & Vidal-Abarca, E. (2010). Understanding and Integrating Multiple Science Texts: Summary Tasks are Sometimes Better Than Argument Tasks. *Reading Psychology*, 31(1), 30–68. doi:10.1080/02702710902733600
- Gillen, J. (2000). Versions of Vygotsky. *British Journal of Educational Studies*, 48(2), 183–198. doi:10.1111/1467-8527.t01-1-00141
- Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6), 511–525. doi:10.1016/j.learninstruc.2006.10.001

- Gluck, J. & Fabrizii, C. (2010). Gender Differences in the Mental Rotations Test Are Partly Explained by Response Format. *Journal of Individual Differences*, 31(2), 106–109. doi:10.1027/1614-0001/a000019
- Gobert, J. D. & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching*, 36(1), 39–53. doi:10.1002/(sici)1098-2736(199901)36:1<39::aid-tea4>3.3.co;2-9
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48(1), 163–189. doi:10.1146/annurev.psych.48.1.163
- Grafton, A. & Rosenberg, D. (2010). *Cartographies of Time: A History of the Timeline*. Princeton Architectural Press.
- Griffin, M. M. & Robinson, D. H. (2005). Does spatial or visual information in maps facilitate text recall? Reconsidering the conjoint retention hypothesis. *Educational Technology Research and Development*, 53(1), 23–36. doi:10.1007/bf02504855
- Gyselinck, V., Meneghetti, C., De Beni, R., & Pazzaglia, F. (2009). The role of working memory in spatial text processing: What benefit of imagery strategy and visuospatial abilities? *Learning and Individual Differences*, 19(1), 12–20. doi:10.1016/j.lindif.2008.08.002
- Haciomeroglu, E. S. & LaVenja, M. (2017). Object-Spatial Imagery and Verbal Cognitive Styles in High School Students. *Perceptual and Motor Skills*, 124(3), 689–702. doi:10.1177/0031512517698555
- Haladjian, H. H. & Mathy, F. (2015). A snapshot is all it takes to encode object locations into spatial memory. *Vision Research*, 107, 133–145. doi:10.1016/j.visres.2014.12.014
- Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words? *Journal of Educational Psychology*, 89(4), 677. doi:10.1037/0022-0663.89.4.677
- Hallam, R. (1972). Thinking and Learning in History. *Teaching*, 8(2), 337–346.

- Harp, S. F. & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414–434. doi:10.1037//0022-0663.90.3.414
- Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183. doi:10.1016/s0166-4115(08)62386-9
- Hart, S. G. & Staveland, L. E. (2005). Results of empirical and theoretical research. In *Ergonomics: Skill, Displays, Controls, and Mental Workload* (Vol. 2, p. 408). Ergonomics: Major Writings. Taylor & Francis.
- Hartmann, M. & Mast, F. W. (2012). Moving along the mental time line influences the processing of future related words. *Consciousness and Cognition*, 21(3), 1558–1562. doi:10.1016/j.concog.2012.06.015
- Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Taylor & Francis.
- Hawking, S. & Penrose, R. (2010). *The Nature of Space and Time*. Princeton Science Library. Princeton University Press.
- Hegarty, M. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32(2), 175–191. doi:10.1016/j.intell.2003.12.001
- Heidig, S., Muller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior*, 44, 81–95. doi:10.1016/j.chb.2014.11.009
- Hicks, J. L., Marsh, R. L., & Russell, E. J. (2000). The properties of retention intervals and their affect on retaining prospective memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1160. doi:10.1037//0278-7393.26.5.1160
- Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*, 69(2), 151–164. doi:10.1016/j.jml.2013.03.002

- Hochpochler, U., Schnotz, W., Rasch, T., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2013). Dynamics of mental model construction from text and graphics. *European Journal of Psychology of Education*, 28(4), 1105–1126.  
doi:10.1007/s10212-012-0156-z
- Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition*, 36(7), 1351–1359. doi:10.3758/MC.36.7.1351
- Hodkinson, A. (2011). Primary History in the 21st century. *The Primary Education Journal of The Historical Association*, (59), 5–6.
- Hoffler, T. N. (2010). Spatial Ability: Its Influence on Learning with Visualizations—a Meta-Analytic Review. *Educational Psychology Review*, 22(3), 245–269.  
doi:10.1007/s10648-010-9126-7
- Hoffler, T. N., Koc-Januchta, M., & Leutner, D. (2017). More Evidence for Three Types of Cognitive Style: Validating the Object-Spatial Imagery and Verbal Questionnaire Using Eye Tracking when Learning with Texts and Pictures: Evidence for three types of cognitive style. *Applied Cognitive Psychology*, 31(1), 109–115.  
doi:10.1002/acp.3300
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.
- Honda, A. & Nihei, Y. (2009). Sex differences in object location memory: The female advantage of immediate detection of changes. *Learning and Individual Differences*, 19(2), 234–237. doi:10.1016/j.lindif.2008.11.005
- Hoodless, P. (1996). *Time and Timelines in the Primary School*. Teaching of History. The Historical Association.
- Hoodless, P. A. (2002). An investigation into children's developing awareness of time and chronology in story. *Journal of Curriculum Studies*, 34(2), 175–177.  
doi:10.1080/00220270110080962
- Houston, J. P. (2014). *Fundamentals of learning and memory*. Academic Press.

- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, 6(6), 435–448.  
doi:10.1038/nrn1684
- Huff, M., Bauhoff, V., & Schwan, S. (2012). Effects of split attention revisited: A new display technology for troubleshooting tasks. *Computers in Human Behavior*, 28(4), 1254–1261. doi:10.1016/j.chb.2012.02.008
- Hugdahl, K., Thomsen, T., & Ersland, L. (2006). Sex differences in visuo-spatial processing: An fMRI study of mental rotation. *Neuropsychologia*, 44(9), 1575–1583.  
doi:10.1016/j.neuropsychologia.2006.01.026
- Huynh, N. T. & Sharpe, B. (2013). An Assessment Instrument to Measure Geospatial Thinking Expertise. *Journal of Geography*, 112(1), 3–17.  
doi:10.1080/00221341.2012.682227
- Inan, F. A., Crooks, S. M., Cheon, J., Ari, F., Flores, R., Kurucay, M., & Paniukov, D. (2015). The reverse modality effect: Examining student learning from interactive computer-based instruction: Multimedia design principles. *British Journal of Educational Technology*, 46(1), 123–130. doi:10.1111/bjet.12129
- Inglot, B., Liu, L., & Antonopoulos, N. (2012). A Framework for Enhanced Timeline Analysis in Digital Forensics. In *2012 IEEE International Conference on Green Computing and Communications* (pp. 253–256). IEEE.  
doi:10.1109/GreenCom.2012.50
- Intraub, H. & Nicklos, S. (1985). Levels of processing and picture memory: The physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(2), 284. doi:10.1037/0278-7393.11.2.284
- Ishihara, M., Keller, P., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: Evidence for the STEARC effect. *Cortex*, 44(4), 454–461.  
doi:10.1016/j.cortex.2007.08.010

- Iversen, J. R. & Balasubramaniam, R. (2016). Synchronization and temporal processing. *Current Opinion in Behavioral Sciences*, 8, 175–180.  
doi:10.1016/j.cobeha.2016.02.027
- Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Applied Psychological Measurement*, 40(8), 559–572. doi:10.1177/01466216166664046
- James, W. (1890). *The Principles of Psychology*. Henry Holt and Co., New York,
- Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior*, 32, 47–53. doi:10.1016/j.chb.2013.11.013
- Jamet, E. & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology*, 32(4), 588–598.  
doi:10.1016/j.cedpsych.2006.07.001
- Johnson, C. I. & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*, 18(2), 178–191. doi:10.1037/a0026923
- Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The Phonological Store of Working Memory: Is It Phonological and Is It a Store? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 656–674. doi:10.1037/0278-7393.30.3.656
- Kalyuga, S. (2007). Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction. *Educational Psychology Review*, 19(4), 509–539.  
doi:10.1007/s10648-007-9054-3
- Kalyuga, S. (2008). Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. *Computers in Human Behavior*, 24(3), 852–861.  
doi:10.1016/j.chb.2007.02.018
- Kalyuga, S. (2013). Effects of Learner Prior Knowledge and Working Memory Limitations on Multimedia Learning. *Procedia - Social and Behavioral Sciences*, 83, 25–29.  
doi:10.1016/j.sbspro.2013.06.005



- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational psychologist*, 38(1), 23–31.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied cognitive psychology*, 13(4), 351–371.  
doi:10.1002/(sici)1099-0720(199908)13:4<351::aid-acp589>3.0.co;2-6
- Kalyuga, S. & Singh, A.-M. (2016). Rethinking the Boundaries of Cognitive Load Theory in Complex Learning. *Educational Psychology Review*, 28(4), 831–852.  
doi:10.1007/s10648-015-9352-0
- Kalyuga, S. & Sweller, J. (2014). The Redundancy Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 247–262). Cambridge Handbooks in Psychology. Cambridge University Press.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4-5), 528–558. doi:10.1080/09541440601056620
- Kang, Y., Harring, J. R., & Li, M. (2015). Reexamining the Impact of Nonnormality in Two-Group Comparison Procedures. *The Journal of Experimental Education*, 83(2), 147–174. doi:10.1080/00220973.2013.876605
- Keating, J. & Sheldon, N. (2011). Government Policy Towards Education and History Teaching 1918 - 39. History in Education Project. Retrieved May 26, 2014, from <http://www.history.ac.uk/history-in-education>
- Keetels, M. & Vroomen, J. (2012). Perception of synchrony between the senses. In *The neural bases of multisensory processes*. Frontiers in Neuroscience. Taylor and Francis.
- Keeting, J. (1976). Government policies towards education 1918-1939.pdf. History in Education Project. Retrieved May 30, 2014, from <http://www.history.ac.uk/history-in-education/project-papers/school-history.html>

- Kelly, J. W. & McNamara, T. P. (2010). Reference frames during the acquisition and development of spatial memories. *Cognition*, 116(3), 409–420.  
doi:10.1016/j.cognition.2010.06.002
- Kendeou, P. & Van Den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & cognition*, 35(7), 1567–1577.
- Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. N. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods*, 13(2), 110–129. doi:10.1037/1082-989X.13.2.110
- Khawaja, M. A., Chen, F., & Marcus, N. (2014). Measuring Cognitive Load Using Linguistic Features: Implications for Usability Evaluation and Adaptive Interaction Design. *International Journal of Human-Computer Interaction*, 30(5), 343–368.  
doi:10.1080/10447318.2013.860579
- Khurana, U., Nguyen, V.-A., Cheng, H.-C., Ahn, J.-w., Chen, X., & Shneiderman, B. (2011). Visual analysis of temporal trends in social networks using edge color coding and metric timelines. In *IEEE International Conference on Social Computing* (pp. 549–554). IEEE. doi:10.1109/PASSAT/SocialCom.2011.212
- Kimble, G. A. (1964). Categories of learning and the problem of definition. In *Categories of human learning* (pp. 34–45). Academic Press.
- Kimble, G. A. (1961). *Hilgard and Marquis' "Conditioning and learning"*. Appleton-Century-Crofts.
- Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. DTIC Document.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49(4), 294. doi:10.1037/0003-066X.49.4.294

- Kintsch, W. & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363. doi:10.1037/0033-295X.85.5.363
- Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education*, 106, 166–171. doi:10.1016/j.compedu.2016.12.006
- Klein Entink, R. H., Kuhn, J.-T., Hornke, L. F., & Fox, J.-P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods*, 14(1), 54–75. doi:10.1037/a0014877
- Korralo, L. (2010a). Middle school children fail to benefit from virtual historical fly-throughs. In *Proceedings of the 7 th Pan-Hellenic Conference with International Participation «ICT in Education*. Korinthos Greece.
- Korralo, L. (2010b). *Use of virtual reality environments to improve the learning of historical chronology* (Doctoral dissertation, Middlesex University, UK).
- Korralo, L., Foreman, N., Boyd-Davis, S., Moar, M., & Coulson, M. (2012). Can multiple "spatial" virtual timelines convey the relatedness of chronological knowledge across parallel domains? *Computers & Education*, 58(2), 856–862. doi:10.1016/j.compedu.2011.10.011
- Koriat, A. & Norman, J. (1985). Mental rotation and visual familiarity. *Perception & Psychophysics*, 37(5), 429–439. doi:10.3758/BF03202874
- Kounios, J. & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 804. doi:10.1037/0278-7393.20.4.804
- Kuhl, T., Scheiter, K., Gerjets, P., & Edelman, J. (2011). The influence of text modality on learning with static and dynamic visualizations. *Computers in Human Behavior*, 27(1), 29–35. doi:10.1016/j.chb.2010.05.008
- Kulhavy, R. W., Lee, J. B., & Caterino, L. C. (1985). Conjoint retention of maps and related discourse. *Contemporary Educational Psychology*, 10(1), 28–37. doi:10.1016/0361-476x(85)90003-7

- Kulhavy, R. W., Stock, W. A., & Kealy, W. A. (1993). How geographic maps increase recall of instructional text. *Educational Technology Research and Development*, 41(4), 47–62. doi:10.1007/BF02297511
- Kulhavy, R. W., Stock, W. A., Woodard, K. A., & Haygood, R. C. (1993). Comparing Elaboration and Dual Coding Theories: The Case of Maps and Text. *The American Journal of Psychology*, 106(4), 483. doi:10.2307/1422965
- Lai, V. & Boroditsky, L. (2013). The immediate and chronic influence of spatio-temporal metaphors on the mental representations of time in English, Mandarin, and Mandarin-English speakers. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00142
- Lakoff, G. & Johnson, M. (2003). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lane, A. R., Ball, K., & Ellison, A. (2015). Dissociating the neural mechanisms of distance and spatial reference frames. *Neuropsychologia*, 74, 42–49. doi:10.1016/j.neuropsychologia.2014.12.019
- Larkin, J. H. & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11(1), 99. doi:10.1111/j.1551-6708.1987.tb00863.x
- Larson-Hall, J. & Herrington, R. (2010). Improving Data Analysis in Second Language Acquisition by Utilizing Modern Developments in Applied Statistics. *Applied Linguistics*, 31(3), 368–390. doi:10.1093/applin/amp038
- Lee, Y.-H., Hsiao, C., & Ho, C.-H. (2014). The effects of various multimedia instructional materials on students' learning responses and outcomes: A comparative experimental study. *Computers in Human Behavior*, 40, 119–132. doi:10.1016/j.chb.2014.07.041
- Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: an aptitude-treatment-interaction study. *Metacognition and Learning*, 11(1), 89–105. doi:10.1007/s11409-015-9149-z

- Lejbak, L., Vrbancic, M., & Crossley, M. (2009). The female advantage in object location memory is robust to verbalizability and mode of presentation of test stimuli. *Brain and Cognition*, 69(1), 148–153. doi:10.1016/j.bandc.2008.06.006
- Leopold, C. & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction*, 22(1), 16–26. doi:10.1016/j.learninstruc.2011.05.005
- Leopold, C. & Mayer, R. E. (2015). An imagination effect in learning from scientific text. *Journal of Educational Psychology*, 107(1), 47–63. doi:10.1037/a0037142
- Leopold, C., Sumfleth, E., & Leutner, D. (2013). Learning with summaries: Effects of representation mode and type of learning activity on comprehension and transfer. *Learning and Instruction*, 27, 40–49. doi:10.1016/j.learninstruc.2013.02.003
- Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4), 1058–1072. doi:10.3758/s13428-013-0334-1
- Leslie, K. C., Low, R., Jin, P., & Sweller, J. (2012). Redundancy and expertise reversal effects when using educational technology to learn primary school science. *Educational Technology Research and Development*, 60(1), 1–13. doi:10.1007/s11423-011-9199-0
- Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning. *Learning and Instruction*, 29, 174–175. doi:10.1016/j.learninstruc.2013.05.004
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. doi:10.1016/j.jesp.2013.03.013
- Libarkin, J. C., Kurdziel, J. P., & Anderson, S. W. (2007). College student conceptions of geological time and the disconnect between ordering and scale. *Journal of Geoscience Education*, 55(5), 413–422.

- Limesurvey. (2015). Stable release. Retrieved August 28, 2017, from  
<https://www.limesurvey.org/stable-release>
- Lin, L., Lee, C. H., Kalyuga, S., Wang, Y., Guan, S., & Wu, H. (2017). The Effect of  
 Learner-Generated Drawing and Imagination in Comprehending a Science Text. *The  
 Journal of Experimental Education*, 85(1), 142–154.  
 doi:10.1080/00220973.2016.1143796
- Lindner, M. A., Eitel, A., Strobel, B., & Koller, O. (2017). Identifying processes underlying  
 the multimedia effect in testing: An eye-movement analysis. *Learning and  
 Instruction*, 47, 91–102. doi:10.1016/j.learninstruc.2016.10.007
- Lindow, S., Fuchs, H. M., Furstenberg, A., Kleber, J., Schweppe, J., & Rummer, R. (2011).  
 On the Robustness of the Modality -Effect: Attempting to Replicate a Basic Finding.  
*Zeitschrift fur Padagogische Psychologie*, 25(4), 231–243.  
 doi:10.1024/1010-0652/a000049
- Linn, M. C. & Petersen, A. C. (1985). Emergence and Characterization of Sex Differences  
 in Spatial Ability: A Meta-Analysis. *Child Development*, 56(6), 1479.  
 doi:10.2307/1130467
- Lippitz, W. (1944). The Child's Understanding of Time. *Phenomenology+ Pedagogy*, 1(2),  
 172–180.
- Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyona, J. (2016). Universality in  
 eye movements and reading: A trilingual investigation. *Cognition*, 147, 1–20.  
 doi:10.1016/j.cognition.2015.10.013
- Loaiza, V. M., McCabe, D. P., Youngblood, J. L., Rose, N. S., & Myerson, J. (2011). The  
 influence of levels of processing on recall from working memory and delayed recall  
 tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5),  
 1258–1263. doi:10.1037/a0023923

- Logie, R. H., Pernet, C. R., Buonocore, A., & Sala, S. D. (2011). Low and high imagers activate networks differentially in mental rotation. *Neuropsychologia*, 49(11), 3071–3077. doi:10.1016/j.neuropsychologia.2011.07.011
- Lourenco, O. (2012). Piaget and Vygotsky: Many resemblances, and a crucial difference. *New Ideas in Psychology*, 30(3), 281–295. doi:10.1016/j.newideapsych.2011.12.006
- Lowe, R. K. (1996). Background knowledge and the construction of a situational representation from a diagram. *European journal of Psychology of Education*, 11(4), 377–397. doi:10.1007/bf03173279
- Lowe, R. & Boucheix, J.-M. (2008). Learning from animated diagrams: How are mental models built? In *International Conference on Theory and Application of Diagrams* (pp. 266–281). Springer.
- Luh, W.-M. & Guo, J.-H. (2007). Approximate sample size formulas for the two-sample trimmed mean test with unequal variances. *British Journal of Mathematical and Statistical Psychology*, 60(1), 137–146. doi:10.1348/000711006X100491
- Lwin, M. O., Morrin, M., & Krishna, A. (2010). Exploring the superadditive effects of scent and pictures on verbal recall: An extension of dual coding theory? *Journal of Consumer Psychology*, 20(3), 317–326. doi:10.1016/j.jcps.2010.04.001
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356. doi:10.1038/nn.3655
- MacLeod, C. M. (1988). Forgotten but not gone: Savings for pictures and words in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 195. doi:10.1037//0278-7393.14.2.195
- Maddison, M. (2011). Developing Pupil's Chronological Understanding: The view from Ofsted. *The Primary Education Journal of The Historical Association*, (59), 8–9.
- Maddison, M. (2014). The National Curriculum for History from September 2014 the view from Ofsted. *Primary History*, 66(Spring 2014), 5–7.

- Maddox, G. B. & Balota, D. A. (2015). Retrieval practice and spacing effects in young and older adults: An examination of the benefits of desirable difficulty. *Memory & Cognition*, 43(5), 760–774. doi:10.3758/s13421-014-0499-6
- Magnani, B. & Musetti, A. (2017). Innate and Cultural Spatial Time: A Developmental Perspective. *Frontiers in Human Neuroscience*, 11. doi:10.3389/fnhum.2017.00215
- Majid, A., Gaby, A., & Boroditsky, L. (2013). Time in terms of space. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00554
- Majid, A., Gaby, A., & Boroditsky, L. (Eds.). (2014). *Time in terms of space*. Frontiers Research Topics. Frontiers Media SA.
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, 57(4), 335–384. doi:10.1016/j.cogpsych.2008.02.004
- Mammarella, N., Fairfield, B., & Di Domenico, A. (2013). When spatial and temporal contiguities help the integration in working memory: "A multimedia learning" approach. *Learning and Individual Differences*, 24, 139–144. doi:10.1016/j.lindif.2012.12.016
- Marin, D., Pitteri, M., Della Puppa, A., Meneghello, F., Biasutti, E., Priftis, K., & Vallesi, A. (2016). Mental time line distortion in right-brain-damaged patients: Evidence from a dynamic spatiotemporal task. *Neuropsychology*, 30(3), 338–345. doi:10.1037/neu0000211
- Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64(1), 17–24. doi:10.1111/j.2044-8295.1973.tb01322.x
- Marks, D. F. (1995). New directions for mental imagery research.
- Mason, L. H., Davison, M. D., Hammer, C. S., Miller, C. A., & Glutting, J. J. (2013). Knowledge, writing, and language outcomes for a reading comprehension and writing intervention. *Reading and Writing*, 26(7), 1133–1158. doi:10.1007/s11145-012-9409-0



- Massa, L. J. & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? *Learning and Individual Differences*, 16(4), 321–335. doi:10.1016/j.lindif.2006.10.001
- Masterman, E. & Rogers, Y. (2002). A framework for designing interactive multimedia to scaffold young children's understanding of historical chronology. *Instructional Science*, 30(3), 221–241. doi:10.1023/A:1015133106888
- Masters, M. S. (1998). The gender difference on the mental rotations test is not due to performance factors. *Memory & Cognition*, 26(3), 444–448. doi:10.3758/BF03201154
- Matlock, T., Ramscar, M., & Boroditsky, L. (2005). On the experiential link between spatial and temporal language. *Cognitive science*, 29(4), 655–664.  
doi:10.1207/s15516709cog0000\_17
- Matthews, W. J. & Meck, W. H. (2014). Time perception: the bad news and the good: Time perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(4), 429–446.  
doi:10.1002/wcs.1298
- Matthews, W. J. & Meck, W. H. (2016). Temporal cognition: Connecting subjective time to perception, attention, and memory. *Psychological Bulletin*, 142(8), 865–907.  
doi:10.1037/bul0000045
- Mayer, R. (1989a). Models for Understanding. *Review of Educational Research*, 59(1), 43–64. doi:10.3102/00346543059001043
- Mayer, R. (2009a). Determining How Multimedia Learning Works. In *Multimedia Learning* (pp. 59–60). Cambridge University Press.
- Mayer, R. (2009b). *Multimedia Learning*. Cambridge University Press.
- Mayer, R. (2009c). The Promise of Multimedia Learning. In *Multimedia Learning* (pp. 19–20). Cambridge University Press.
- Mayer, R. (2014a). *The Cambridge Handbook of Multimedia Learning*. Cambridge Handbooks in Psychology. Cambridge University Press.

- Mayer, R. E. (1989b). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81(2), 240. doi:10.1037//0022-0663.81.2.240
- Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, 8(4), 357–371. doi:10.1007/BF01463939
- Mayer, R. E. (1999). Multimedia aids to problem-solving transfer. *International Journal of Educational Research*, 31(7), 611–623. doi:10.1016/S0883-0355(99)00027-0
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press.
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2), 167–171.  
doi:10.1016/j.learninstruc.2009.02.012
- Mayer, R. E. (2014b). Incorporating motivation into multimedia learning. *Learning and Instruction*, 29, 171–173. doi:10.1016/j.learninstruc.2013.04.003
- Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, 88(1), 64. doi:10.1037/0022-0663.88.1.64
- Mayer, R. E. & Fiorella, L. (2014). 12 Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In *The Cambridge handbook of multimedia learning* (p. 279). Cambridge University Press.
- Mayer, R. E. & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82(4), 715. doi:10.1037/0022-0663.82.4.715
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93(1), 187–198. doi:10.1037//0022-0663.93.1.187

- Mayer, R. E. & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology*, 100(2), 380–386.  
doi:10.1037/0022-0663.100.2.380
- Mayer, R. E. & Massa, L. J. (2003). Three Facets of Visual and Verbal Learners: Cognitive Ability, Cognitive Style, and Learning Preference. *Journal of Educational Psychology*, 95(4), 833–846. doi:10.1037/0022-0663.95.4.833
- Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312. doi:10.1037//0022-0663.90.2.312
- Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86(3), 389. doi:10.1037/0022-0663.86.3.389
- Mayes, D. K., Sims, V. K., & Koonce, J. M. (2001). Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics*, 28(6), 367–378. doi:10.1016/s0169-8141(01)00043-9
- McCrink, K. & Opfer, J. E. (2014). Development of spatial-numerical associations. *Current Directions in Psychological Science*, 23(6), 439–445.  
doi:10.1177/0963721414549751
- McCrudden, M. T., Hushman, C. J., & Marley, S. C. (2014). Exploring the Boundary Conditions of the Redundancy Principle. *The Journal of Experimental Education*, 82(4), 537–554. doi:10.1080/00220973.2013.813368
- McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational psychologist*, 35(1), 13–23.  
doi:10.1207/S15326985EP3501\_3
- McGlone, M. S. & Harding, J. L. (1998). Back (or forward?) to the future: The role of perspective in temporal language comprehension. *Journal of Experimental*

- Psychology-Learning Memory and Cognition*, 24(5), 1211–1223.  
doi:10.1037/0278-7393.24.5.1211
- McNamara, D. S. & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247–288.  
doi:10.1080/01638539609544975
- McNeill, A. L., Doolittle, P. E., & Hicks, D. (2009). The effects of training, modality, and redundancy on the development of a historical inquiry strategy in a multimedia learning environment. *Journal of Interactive Online Learning*, 8(3), 255–269.
- Meneghetti, C., De Beni, R., Pazzaglia, F., & Gyselinck, V. (2011). The role of visuo-spatial abilities in recall of spatial descriptions: A mediation model. *Learning and Individual Differences*, 21(6), 719–723. doi:10.1016/j.lindif.2011.07.015
- Meneghetti, C., Fiore, F., Borella, E., & De Beni, R. (2011). Learning a map of environment: The role of visuo-spatial abilities in young and older adults. *Applied Cognitive Psychology*, 25(6), 952–959. doi:10.1002/acp.1788
- Meneghetti, C., Gyselinck, V., Pazzaglia, F., & De Beni, R. (2009). Individual differences in spatial text processing: High spatial ability can compensate for spatial working memory interference. *Learning and Individual Differences*, 19(4), 577–589.  
doi:10.1016/j.lindif.2009.07.007
- Meneghetti, C., Labate, E., Grassano, M., Ronconi, L., & Pazzaglia, F. (2014). The role of visuospatial and verbal abilities, styles and strategies in predicting visuospatial description accuracy. *Learning and Individual Differences*, 36, 117–123.  
doi:10.1016/j.lindif.2014.10.019
- Merchant, H. & Yarrow, K. (2016). How the motor system both encodes and influences our sense of time. *Current Opinion in Behavioral Sciences*, 8, 22–27.  
doi:10.1016/j.cobeha.2016.01.006

- Merkt, M. & Schwan, S. (2017). What you see is what you remember? Depictions of historical figures influence memory for historical facts. *Learning and Instruction*. doi:10.1016/j.learninstruc.2017.05.004
- Merritt, D. J., Casasanto, D., & Brannon, E. M. (2010). Do monkeys think in metaphors? Representations of space and time in monkeys and humans. *Cognition*, 117(2), 191–202. doi:10.1016/j.cognition.2010.08.011
- Miles, L. K., Betka, E., Pendry, L. F., & Macrae, C. N. (2010). Mapping temporal constructs: Actions reveal that time is a place. *The Quarterly Journal of Experimental Psychology*, 63(11), 2113–2119. doi:10.1080/17470218.2010.524932
- Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, 63, 81–97. doi:10.1037/0033-295x.101.2.343
- Mioni, G., Stablum, F., & Grondin, S. (2014). Interval discrimination across different duration ranges with a look at spatial compatibility and context effects. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.00717
- Mitchell, M. (2014). Fitting issues: The visual representation of time in family tree diagrams. *Sign Systems Studies*, 42(2-3), 241. doi:10.12697/SSS.2014.42.2-3.05
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130(4), 621–640. doi:10.1037//0096-3445.130.4.621
- Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, 22(3), 149–158. doi:10.1111/j.1365-2729.2006.00170.x
- Moreno, R. & Mayer, R. E. (1999a). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(2), 358. doi:10.1037/0022-0663.91.2.358

- Moreno, R. & Mayer, R. E. (1999b). Multimedia-supported metaphors for meaning making in mathematics. *Cognition and instruction*, 17(3), 215–248.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, 16(5), 519–533.
- Morrison, J., Watson, G. S., & Morrison, G. R. (2015). Exploring the redundancy effect in print-based instruction containing representations: Redundancy effect. *British Journal of Educational Technology*, 46(2), 423–436. doi:10.1111/bjet.12140
- Morrone, M. C., Cicchini, M., & Burr, D. C. (2010). Spatial maps for time and motion. *Experimental Brain Research*, 206(2), 121–128. doi:10.1007/s00221-010-2334-z
- Mulder, L. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological psychology*, 34(2), 205–236.  
doi:10.1016/0301-0511(92)90016-N
- Muller, D. A., Lee, K. J., & Sharma, M. D. (2008). Coherence or interest: Which is most important in online multimedia learning? *Australasian Journal of Educational Technology*, 24(2). doi:10.14742/ajet.1223
- Mulligan, N. W. & Picklesimer, M. (2012). Levels of processing and the cue-dependent nature of recollection. *Journal of Memory and Language*, 66(1), 79–92.  
doi:10.1016/j.jml.2011.10.001
- Murai, Y., Whitaker, D., & Yotsumoto, Y. (2016). The centralized and distributed nature of adaptation-induced misjudgments of time. *Current Opinion in Behavioral Sciences*, 8, 117–123. doi:10.1016/j.cobeha.2016.02.011
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53(1), 55–81.  
doi:10.1146/annurev.psych.53.100901.135131

- Nesbit, J. C. & Adesope, O. O. (2006). Learning With Concept and Knowledge Maps: A Meta-Analysis. *Review of Educational Research*, 76(3), 413–448.  
doi:10.3102/00346543076003413
- Newbern, D., Dansereau, D. F., & Patterson, M. E. (1997). Spatial-semantic display processing: The role of spatial structure on recall. *Contemporary Educational Psychology*, 22(3), 319–337. doi:10.1006/ceps.1997.0936
- Nguyen, P. H., Xu, K., Walker, R., & Wong, B. W. (2014). SchemaLine: Timeline Visualization for Sensemaking. (pp. 225–233). IEEE. doi:10.1109/IV.2014.14
- Nichols, H. (1891). The Psychology of Time. *The American Journal of Psychology*, 4(1), 60. doi:10.2307/1411839
- Nieder, A. & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37(1), 149–157.
- Nixon, L., Gregson, M., & Spedding, T. (2007). Pedagogy and the intuitive appeal of learning styles in post-compulsory education in England. *Journal of Vocational Education & Training*, 59(1), 39–51. doi:10.1080/13636820601144088
- Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (pp. 420–423). ACM.  
doi:10.1145/2414536.2414602
- Nuerk, H.-C., Wood, G., & Willmes, K. (2005). The Universal SNARC Effect: The Association between Number Magnitude and Space is Amodal. *Experimental Psychology*, 52(3), 187–194. doi:10.1027/1618-3169.52.3.187
- Nunez, R. E. (2011). No Innate Number Line in the Human Brain. *Journal of Cross-Cultural Psychology*, 42(4), 651–668. doi:10.1177/0022022111406097
- Nunez, R. & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, 17(5), 220–229. doi:10.1016/j.tics.2013.03.008

- Oberfoell, A. & Correia, A. (2016). Understanding the role of the modality principle in multimedia learning environments: Understanding the modality principle. *Journal of Computer Assisted Learning*, 32(6), 607–617. doi:10.1111/jcal.12157
- Ofsted. (2007). *History in the balance*.
- Ofsted. (2011). History for All.
- Olson, M. (2015). *An Introduction to Theories of Learning: Ninth Edition*. Taylor & Francis.
- Olsson, J. & Boldt, M. (2009). Computer forensic timeline visualization tool. *Digital Investigation*, 6, S78–S87. doi:10.1016/j.diin.2009.06.008
- Oxford English Dictionary. (2012). Timeline - definition of timeline in English | Oxford Dictionaries. Retrieved August 17, 2017, from <https://en-oxforddictionaries-com.ezproxy1.bath.ac.uk/definition/timeline>
- Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, 19(3), 228–242. doi:10.1016/j.learninstruc.2008.04.003
- Paas, F. G. & Van Merriënboer, J. J. (1993). The Efficiency of Instructional Conditions An Approach to Combine Mental Effort and Performance Measures. *Human Factors*, 35(4), 737–743. doi:10.1177/001872089303500412
- Paas, F. G. & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational psychology review*, 6(4), 351–371.
- Paas, F. G., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79(1), 419–430. doi:10.2466/pms.1994.79.1.419
- Paas, F., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning. In *Recent Innovations in Educational Technology That Facilitate Student Learning*, (pp. 11–35). Charlotte, NC: Information Age Publishing Inc.



- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory : Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, 32(1), 1–8. doi:10.1023/B:TRUC.0000021806.17516.d0
- Paas, F. & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning*, 27, 27–42.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63–71. doi:10.1207/S15326985EP3801\_8
- Paivio, A. (2008). Dual coding theory and education. In *Educating the other America* (pp. 227–242). Paul H. Brookes Publishing Company.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 45(3), 255. doi:10.1037/h0084295
- Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of Kousta et al. (2011). *Journal of Experimental Psychology: General*, 142(1), 282–287. doi:10.1037/a0027004
- Paivio, A. (2014). Intelligence, dual coding theory, and the brain. *Intelligence*, 47, 141–158. doi:10.1016/j.intell.2014.09.002
- Paivio, A., Clark, J. M., & Lambert, W. E. (1988). Bilingual dual-coding theory and semantic repetition effects on recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(1), 163. doi:10.1037/0278-7393.14.1.163
- Paivio, A. & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, 5(2), 176–206. doi:10.1016/0010-0285(73)90032-7
- Paivio, A. & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 532–539. doi:10.1016/S0022-5371(81)90156-0
- Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11(4), 137–138.

- Paivio, A. & Sadoski, M. (2011). Lexicons, Contexts, Events, and Images: Commentary on Elman (2009) From the Perspective of Dual Coding Theory. *Cognitive Science*, 35(1), 198–209. doi:10.1111/j.1551-6709.2010.01146.x
- Palmer, D. (2005). A Motivational View of Constructivist-informed Teaching. *International Journal of Science Education*, 27(15), 1853–1881. doi:10.1080/09500690500339654
- Park, B., Flowerday, T., & Brunken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior*, 44, 267–278. doi:10.1016/j.chb.2014.10.061
- Park, B., Plass, J. L., & Brunken, R. (2014). Cognitive and affective processes in multimedia learning. *Learning and Instruction*, 29, 125–127. doi:10.1016/j.learninstruc.2013.05.005
- Parsons, L. M. (1987). Visual discrimination of abstract mirror-reflected three-dimensional objects at many orientations. *Attention, Perception, & Psychophysics*, 42(1), 49–59. doi:10.3758/BF03211513
- Parsons, T. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia*, 42(4), 555–562. doi:10.1016/j.neuropsychologia.2003.08.014
- Paschke, K., Jordan, K., Wustenberg, T., Baudewig, J., & Leo Muller, J. (2012). Mirrored or identical – Is the role of visual perception underestimated in the mental rotation process of 3d-objects?: A combined fMRI-eye tracking-study. *Neuropsychologia*, 50(8), 1844–1851. doi:10.1016/j.neuropsychologia.2012.04.010
- Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007-2004. *National Center for Education Research*.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles concepts and evidence. *Psychological Science in the Public Interest*, 9(3), 105–119. doi:10.1111/j.1539-6053.2009.01038.x

- Pasqualotto, A., Spiller, M. J., Jansari, A. S., & Proulx, M. J. (2013). Visual experience facilitates allocentric spatial representation. *Behavioural Brain Research*, 236, 175–179. doi:10.1016/j.bbr.2012.08.042
- Pazzaglia, F. & Moe, A. (2013). Cognitive styles and mental rotation ability in map learning. *Cognitive Processing*, 14(4), 391–399. doi:10.1007/s10339-013-0572-2
- Peebles, D. & Jones, C. (2014). A model of object location memory. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36). Quebec City, Canada: Curran Associates, Inc.
- Piaget, J. (1969). Conclusions (A. Pomerans, Trans.). In, *The Child's Conception of Time* (pp. 255–256). Routledge and Kegan Paul.
- Pickford, T. (2011). Timelines and technology. *The Primary Education Journal of The Historical Association*, (59), 25–26.
- Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, 19(3), 291–312. doi:10.1002/acp.1086
- Pitt, B. & Casasanto, D. (2016). Reading experience shapes the mental timeline but not the mental number line. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2753e2758). Austin, TX: Cognitive Science Society.
- Plaisant, C., Milash, B., Rose, A., Widoff, S., & Shneiderman, B. (1996). LifeLines: visualizing personal histories. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 221–227). ACM. doi:10.1145/238386.238493
- Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, 19(2), 221–243. doi:10.1016/S0747-5632(02)00015-8
- Ponce, H. R. & Mayer, R. E. (2014). An eye movement analysis of highlighting and graphic organizer study aids for learning from expository text. *Computers in Human Behavior*, 41, 21–32. doi:10.1016/j.chb.2014.09.010

- Poppel, E. (2004). Lost in time: a historical frame, elementary processing units and the 3-second window. *Acta neurobiologiae experimentalis*, 64(3), 295–302.
- Prangmsa, M. E. (2007). *Multimodal representations in collaborative history learning* (Doctoral dissertation, University of Utrecht).
- Prangmsa, M. E., Boxtel, C. A. M., & Kanselaar, G. (2008). Developing a 'big picture': Effects of collaborative construction of multimodal representations in history. *Instructional Science*, 36(2), 117–136. doi:10.1007/s11251-007-9026-5
- Prangmsa, M. E., Boxtel, C. A. M., Kanselaar, G., & Kirschner, P. A. (2009). Concrete and abstract visualizations in history learning tasks. *British Journal of Educational Psychology*, 79(2), 371–387. doi:10.1348/000709908X379341
- Provost, A. & Heathcote, A. (2015). Titrating decision processes in the mental rotation task. *Psychological Review*, 122(4), 735–754. doi:10.1037/a0039706
- Rau, C. (1953). Theories of Time in Ancient Philosophy. *The Philosophical Review*, 62(4), 514. doi:10.2307/2182458
- Rayner, K. (2015). Eye Movements in Reading. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 631–634). Elsevier. doi:10.1016/B978-0-08-097086-8.54008-2
- Razali, N. M. & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33. doi:10.1515/bile-2015-0008
- Rey, G. D. (2014). Seductive details and attention distraction - An eye tracker experiment. *Computers in Human Behavior*, 32, 133–144. doi:10.1016/j.chb.2013.11.017
- Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19–36. doi:10.1016/j.edurev.2015.12.003

- Robert, M. & Savoie, N. (2006). Are there gender differences in verbal and visuospatial working-memory resources? *European Journal of Cognitive Psychology*, 18(3), 378–397. doi:10.1080/09541440500234104
- Robinson, D. H., Robinson, S. L., & Katayama, A. D. (1999). When words are represented in memory like pictures Evidence for spatial encoding of study materials. *Contemporary Education Psychology*, 24, 38–54. doi:10.1006/ceps.1998.0979
- Roediger, H. L. & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. doi:10.1016/j.tics.2010.09.003
- Roediger, H. L. & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17(3), 249–255. doi:10.1111/j.1467-9280.2006.01693.x
- Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2015). Matching learning style to instructional method: Effects on comprehension. *Journal of Educational Psychology*, 107(1), 64–78. doi:10.1037/a0037478
- Rohde, M. & Ernst, M. O. (2014). Twists and Turns in Visuomotor Time Perception and Temporal Recalibration. *Procedia - Social and Behavioral Sciences*, 126, 69–70. doi:10.1016/j.sbspro.2014.02.319
- Roitman, J. D., Brannon, E. M., Andrews, J. R., & Platt, M. L. (2007). Nonverbal representation of time and number in adults. *Acta Psychologica*, 124(3), 296–318. doi:10.1016/j.actpsy.2006.03.008
- Rop, G., Verkoeijen, P. P. J. L., & van Gog, T. (2017). With task experience students learn to ignore the content, not just the location of irrelevant information. *Journal of Cognitive Psychology*, 29(5), 599–606. doi:10.1080/20445911.2017.1299154
- Rosenberg, D. & Grafton, A. (2013). *Cartographies of Time: A History of the Timeline*. Princeton Architectural Press.

- Ross, S. M. & Morrison, G. R. (1989). In search of a happy medium in instructional technology research: Issues concerning external validity, media replications, and learner control. *Educational Technology Research and Development*, 37(1), 19–33.
- Ross, S. M., Morrison, G. R., & Lowther, D. L. (2010). Educational Technology Research Past and Present: Balancing: Rigor and Relevance to Impact School Learning. *Contemporary Educational Technology*, 1(1).
- Rubin, D. C. & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103(4), 734.  
doi:10.1037/0033-295X.103.4.734
- Rugani, R., Lunghi, M., Di Giorgio, E., Regolin, L., Dalla Barba, B., Vallortigara, G., & Simion, F. (2017). A mental number line in human newborns. *bioRxiv*, 159335.  
doi:10.1101/159335
- Sadoski, M., Kealy, W. A., Goetz, E. T., & Paivio, A. (1997). Concreteness and imagery effects in the written composition of definitions. *Journal of Educational Psychology*, 89(3), 518–526. doi:10.1037/0022-0663.89.3.518
- Sadoski, M. & Paivio, A. (2004). A Dual Coding Theoretical Model of Reading. In *Theoretical models and processes of reading* (5th ed., pp. 1329–1362). International Reading Association.
- Sadoski, M., Paivio, A., & Goetz, E. T. (1991). Commentary: A Critique of Schema Theory in Reading and a Dual Coding Alternative. *Reading Research Quarterly*, 26(4), 463.  
doi:10.2307/747898
- Salomon, G. (2012). *Interaction of Media Cognition and Learning: An Exploration of How Symbolic Forms Cultivate Mental Skills and Affect Knowledge Acquisition*. Taylor & Francis.
- Salthouse, T. A. (1982). *Adult Cognition* (C. J. Brainerd, Ed.). Springer Series in Cognitive Development. New York, NY: Springer New York.

- Salthouse, T. A., Babcock, R. L., Skovronek, E., Mitchell, D. R., & Palmon, R. (1990). Age and experience effects in spatial visualization. *Developmental Psychology*, 26(1), 128. doi:10.1037/0012-1649.26.1.128
- Sanchez, C. A. & Branaghan, R. J. (2009). The interaction of map resolution and spatial abilities on map learning. *International Journal of Human-Computer Studies*, 67(5), 475–481. doi:10.1016/j.ijhcs.2008.12.003
- Sanchez, C. A. & Wiley, J. (2014). The role of dynamic spatial ability in geoscience text comprehension. *Learning and Instruction*, 31, 33–45. doi:10.1016/j.learninstruc.2013.12.007
- Scevak, J. J. & Moore, P. J. (1998). Levels of Processing Effects on Learning from Texts with Maps. *Educational Psychology*, 18(2), 133–155. doi:10.1080/0144341980180201
- Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology*, 39(4), 275–286. doi:10.1016/j.cedpsych.2014.07.003
- Schnotz, W. & Baadte, C. (2015). Surface and deep structures in graphics comprehension. *Memory & Cognition*, 43(4), 605–618. doi:10.3758/s13421-014-0490-2
- Schnotz, W. & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. doi:10.1016/S0959-4752(02)00017-8
- Schnotz, W. & Kurschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, 19(4), 469–508. doi:10.1007/s10648-007-9053-4
- Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *Journal of Educational Psychology*, 106(4), 974–989. doi:10.1037/a0037054

- Schnotz, W., Mengelkamp, C., Baadte, C., & Hauck, G. (2014). Focus of attention and choice of text modality in multimedia learning. *European Journal of Psychology of Education*, 29(3), 483–501. doi:10.1007/s10212-013-0209-y
- Schrivver, K. A. (1989). Evaluating text quality: The continuum from text-focused to reader-focused methods. *IEEE Transactions on Professional Communication*, 32(4), 238–255. doi:10.1109/47.44536
- Schrivver, K. A. (2000). Readability formulas in the new millennium: what's the use? *ACM Journal of Computer Documentation (JCD)*, 24(3), 138–140. doi:10.1145/344599.344638
- Schuler, A., Arndt, J., & Scheiter, K. (2015). Processing multimedia material: Does integration of text and pictures result in a single or two interconnected mental representations? *Learning and Instruction*, 35, 62–72. doi:10.1016/j.learninstruc.2014.09.005
- Schuler, A., Scheiter, K., & Gerjets, P. (2010). Can text content influence the effectiveness of diagrams? In *Diagrammatic Representation and Inference* (pp. 322–324). Springer.
- Schuler, A., Scheiter, K., & Gerjets, P. (2013). Is spoken text always better? Investigating the modality and redundancy effect with longer text presentation. *Computers in Human Behavior*, 29(4), 1590–1601. doi:10.1016/j.chb.2013.01.047
- Schuler, A., Scheiter, K., Rummer, R., & Gerjets, P. (2012). Explaining the modality effect in multimedia learning: Is it due to a lack of temporal contiguity with written text and pictures? *Learning and Instruction*, 22(2), 92–102. doi:10.1016/j.learninstruc.2011.08.001
- Schwaborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology*, 102(4), 872–879. doi:10.1037/a0019640
- Schweppe, J., Eitel, A., & Rummer, R. (2015). The multimedia effect and its stability over time. *Learning and Instruction*, 38, 24–33. doi:10.1016/j.learninstruc.2015.03.001



- Schweppe, J. & Rummer, R. (2012). Long-term multimedia learning. *Staging Knowledge and Experience: How to Take Advantage of Representational Technologies in Education and Training*, 196–198. doi:10.1016/j.chb.2016.02.035
- Schweppe, J. & Rummer, R. (2016). Integrating written text and graphics as a desirable difficulty in long-term multimedia learning. *Computers in Human Behavior*, 60, 131–137. doi:10.1016/j.chb.2016.02.035
- Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, 13(2), 227–237. doi:10.1016/S0959-4752(02)00022-1
- Shaki, S., Fischer, M. H., & Gobel, S. M. (2012). Direction counts: A comparative study of spatially directional counting biases in cultures with different reading directions. *Journal of Experimental Child Psychology*, 112(2), 275–281. doi:10.1016/j.jecp.2011.12.005
- Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, 16(2), 328–331. doi:10.3758/PBR.16.2.328
- Shayer, M. (2003). Not just Piaget; not just Vygotsky, and certainly not Vygotsky as alternative to Piaget. *Learning and Instruction*, 13(5), 465–485. doi:10.1016/S0959-4752(03)00092-6
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1), 156–163. doi:10.1016/S0022-5371(67)80067-7
- Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703. doi:10.1126/science.171.3972.701
- Shepard, S. & Metzler, D. (1988). Mental rotation: effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 3. doi:10.1037//0096-1523.14.1.3

- Silverman, I., Choi, J., & Peters, M. (2007). The Hunter-Gatherer Theory of Sex Differences in Spatial Abilities: Data from 40 Countries. *Archives of Sexual Behavior*, 36(2), 261–268. doi:10.1007/s10508-006-9168-6
- Silverman, I. & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind* (pp. 533–549). Oxford University Press.
- Simon, H. A. (1974). How big is a chunk? *Science*, 183(February), 482–488. doi:10.1126/science.183.4124.482
- Simsek, A. (2007). The Improvement of Chronological Perceptions among Fifth Grade Students: A Quasi-Experimental Study. *Educational Sciences: Theory & Practice*, 7(1).
- Sinha, C. & Gardenfors, P. (2014). Time, space, and events in language and cognition: a comparative view: Time, space, and events in language and cognition. *Annals of the New York Academy of Sciences*, 1326(1), 72–81. doi:10.1111/nyas.12491
- Slavin, R. E. (2008). Perspectives on Evidence-Based Research in Education—What Works? Issues in Synthesizing Educational Program Evaluations. *Educational Researcher*, 37(1), 5–14. doi:10.3102/0013189X08314117
- Soderstrom, N. C. & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, 73, 99–115. doi:10.1016/j.jml.2014.03.003
- Spanjers, I. A., Wouters, P., van Gog, T., & van Merriënboer, J. J. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior*, 27(1), 46–52. doi:10.1016/j.chb.2010.05.011
- Spichtig, A. N., Hiebert, E. H., Vorstius, C., Pascoe, J. P., David Pearson, P., & Radach, R. (2016). The Decline of Comprehension-Based Silent Reading Efficiency in the United States: A Comparison of Current Data With Performance in 1960. *Reading Research Quarterly*, 51(2), 239–259. doi:10.1002/rrq.137

- Spirgel, A. S. & Delaney, P. F. (2016). Does Writing Summaries Improve Memory for Text? *Educational Psychology Review*, 28(1), 171–196. doi:10.1007/s10648-014-9290-2
- Srinivasan, M. & Carey, S. (2010). The long and the short of it: On the nature and origin of functional overlap between representations of space and time. *Cognition*, 116(2), 217–241. doi:10.1016/j.cognition.2010.05.005
- Stow, W. & Haydn, T. (2000a). Issues in History Teaching. In J. Arthur & R. Philips (Eds.), *Issues in History Teaching* (p. 92). Issues in History Teaching. Routledge.
- Stow, W. & Haydn, T. (2000b). Issues in the teaching of chronology. In J. Arthur & R. Phillips (Eds.), *Issues in History Teaching* (pp. 83–97). Issues in History Teaching. London: Routledge.
- Sutton Trust. (2018). Teaching and Learning Toolkit. Retrieved March 24, 2018, from <https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 261–262. doi:10.1207/s15516709cog1202\_4
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4, 295–312. doi:10.1016/0959-4752(94)90003-5
- Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, 22(2), 123–138. doi:10.1007/s10648-010-9128-5
- Sweller, J. (2015). Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition*. doi:10.1016/j.jarmac.2015.12.002
- Sweller, J., Ayres, P., & Mayer, R. E. (2006). The split-attention principle in multimedia learning. In *Cambridge handbook of multimedia learning* (pp. 135–146).
- Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 185–233. doi:10.1207/s1532690xc1203\_1

- Tabbers, H. K., Martens, R. L., & Merrienboer, J. J. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, 74(1), 71–81. doi:10.1348/000709904322848824
- Tagaris, G. A., Kim, S.-G., Strupp, J. P., Andersen, P., Ugurbil, K., & Georgopoulos, A. P. (1997). Mental rotation studied by functional magnetic resonance imaging at high field (4 Tesla): Performance and cortical activation. *Journal of Cognitive Neuroscience*, 9(4), 419–432. doi:10.1162/jocn.1997.9.4.419
- Takahashi, T. (2006). Time-estimation error following Weber-Fechner law may explain subadditive time-discounting. *Medical Hypotheses*, 67(6), 1372–1374. doi:10.1016/j.mehy.2006.05.056
- Tarr, M. (2015). Shepard\_metzler.zip. Retrieved August 3, 2015, from [http://wiki.cmb.cmu.edu/images/Shepard\\_Metzler.zip](http://wiki.cmb.cmu.edu/images/Shepard_Metzler.zip)
- Tarr, M. J. & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21(2), 233–282. doi:10.1016/0010-0285(89)90009-1
- Thode, H. (2002). *Testing For Normality*. Statistics, textbooks and monographs. Boca Raton: Taylor & Francis.
- Thomas, N. (1990). *Primary Education from Plowden to the 1990s*. Education Policy Perspectives. Falmer Press.
- Thorndyke, P. W. & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology*, 12, 137–175. doi:10.1016/0010-0285(80)90006-7
- Thornton, S. & Vukelich, R. (1988). Effects of Children's Understanding of Time Concepts on Historical Understanding. *Theory and Research in Social Education*, 16(1), 68–82. doi:10.1080/00933104.1988.10505556

- Torralbo, A., Santiago, J., & Lupianez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cognitive Science*, 30(4), 745–757.  
doi:10.1207/s15516709cog0000\_67
- Treude, C. & Storey, M.-A. (2009). ConcernLines: A timeline view of co-occurring concerns. In *Proceedings of the 31st International Conference on Software Engineering* (pp. 575–578). IEEE Computer Society. doi:10.1109/icse.2009.5070559
- Ungar, S., Blades, M., & Spencer, C. (2000). Can a tactile map facilitate learning of related information by blind and visually impaired people? a test of the conjoint retention hypothesis. *Proceedings of Thinking with Diagrams*, 98.
- Unsworth, N. & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133(6), 1038–1066.  
doi:10.1037/0033-2909.133.6.1038
- Vallesi, A., Binns, M. A., & Shallice, T. (2008). An effect of spatial-temporal association of response codes: Understanding the cognitive representations of time. *Cognition*, 107(2), 501–527. doi:10.1016/j.cognition.2007.10.011
- Vallesi, A., McIntosh, A. R., & Stuss, D. T. (2011). How time modulates spatial responses. *Cortex*, 47(2), 148–156. doi:10.1016/j.cortex.2009.09.005
- van den Broek, G. S. E., Segers, E., Takashima, A., & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, 22(7), 803–812. doi:10.1080/09658211.2013.831455
- Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, 31(2), 142–166. doi:10.1016/j.cedpsych.2005.04.001
- Van Meter, P. & Garner, J. (2005). The Promise and Practice of Learner-Generated Drawing: Literature Review and Synthesis. *Educational Psychology Review*, 17(4), 285–325. doi:10.1007/s10648-005-8136-3

- Van Zandt, T. & Townsend, J. T. (2014). Designs for and analyses of response time experiments. *The oxford handbook of quantitative methods: Foundations, 1*, 260. doi:10.1093/oxfordhb/9780199934874.013.0014
- van Genuchten, E., Scheiter, K., & Schuler, A. (2012). Examining learning from text and pictures for different task types: Does the multimedia effect differ for conceptual, causal, and procedural tasks? *Computers in Human Behavior*, 28(6), 2209–2218. doi:10.1016/j.chb.2012.06.028
- van Gog, T. (2014). 11 The Signaling (or Cueing) Principle in Multimedia Learning. *The Cambridge handbook of multimedia learning*, 263.
- van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and Frequency of Mental Effort Measurement: Evidence in Favour of Repeated Measures: Measuring mental effort. *Applied Cognitive Psychology*, 26(6), 833–839. doi:10.1002/acp.2883
- van Gog, T. & Paas, F. (2008). Instructional Efficiency: Revisiting the Original Construct in Educational Research. *Educational Psychologist*, 43(1), 16–26. doi:10.1080/00461520701756248
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. doi:10.1080/17470218.2013.850521
- van Merriënboer, J. J. G. & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), 147–177. doi:10.1007/s10648-005-3951-0
- Vekiri, I. (2002). What is the value of graphical displays in learning? *Educational Psychology Review*, 14(3), 261–312. doi:10.1023/A:1016064429161
- Verdi, M. P., Johnson, J. T., Stock, W. A., Kulhavy, R. W., & Whitman-Ahern, P. (1997). Organized Spatial Displays and Texts: Effects of Presentation Order and Display Type

- on Learning Outcomes. *The Journal of Experimental Education*, 65(4), 303–317.  
doi:10.1080/00220973.1997.10806606
- Verdi, M. P. & Kulhavy, R. W. (2002). Learning with maps and texts: An overview.  
*Educational Psychology Review*, 14(1), 27–46.
- Verdi, M. P., Stamm, J., Johnson, J. T., & Jamison, M. (2001). Map Edges: Focal Points for Facilitating Text Recall. *Contemporary Educational Psychology*, 26(2), 211–226.  
doi:10.1006/ceps.2000.1047
- Viarouge, A., Hubbard, E. M., & McCandliss, B. D. (2014). The Cognitive Mechanisms of the SNARC Effect: An Individual Differences Approach. *PLOS ONE*, 9(4), e95756.  
doi:10.1371/journal.pone.0095756
- Voss, J. F. & Silfies, L. (1996). Learning From History Text The Interaction of Knowledge and Comprehension Skill with Text Structure. *Cognition and Instruction*, 14(1), 45–68. doi:10.1207/s1532690xc1401\_2
- Vroomen, J. & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics*, 72(4), 871–884. doi:10.3758/app.72.4.871
- Walker, E. J., Bergen, B. K., & Nunez, R. (2014). Disentangling Spatial Metaphors for Time Using Non-spatial Responses and Auditory Stimuli. *Metaphor and Symbol*, 29(4), 316–327. doi:10.1080/10926488.2014.948801
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(11), 483–488.  
doi:10.1016/j.tics.2003.09.002
- Weger, U. W. & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review*, 15(2), 426–430.  
doi:10.3758/PBR.15.2.426
- Welcome, S. E., Paivio, A., McRae, K., & Joanisse, M. F. (2011). An electrophysiological study of task demands on concreteness effects: evidence for dual coding theory. *Experimental Brain Research*, 212(3), 347–358. doi:10.1007/s00221-011-2734-8

- Westelinck, K. D., Valcke, M., De Craene, B., & Kirschner, P. (2005). Multimedia learning in social sciences: limitations of external graphical representations. *Computers in Human Behavior*, 21(4), 555–573. doi:10.1016/j.chb.2004.10.030
- Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11(6), 571–580. doi:10.1080/09658210244000414
- Whelan, R. (2010). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 9. doi:10.1007/BF03395630
- Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). The development of the picture-superiority effect. *British Journal of Developmental Psychology*, 24(4), 767–773. doi:10.1348/026151005X74153
- Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior*, 26(3), 474–481. doi:10.1016/j.chb.2009.12.006
- Wilcox, R. R. (1992). Why Can Methods for Comparing Means Have Relatively Low Power, and What Can You Do to Correct the Problem? *Current Directions in Psychological Science*, 1(3), 101–105. doi:10.1111/1467-8721.ep10768801
- Wilcox, R. (2016). *Introduction to Robust Estimation and Hypothesis Testing*. Statistical Modeling and Decision Science. Elsevier Science. doi:10.1016/c2010-0-67044-1
- Wilkening, J. & Fabrikant, S. (2011). The Effect of Gender and Spatial Abilities on Map Use Preferences and Performance in Road Selection Tasks. In *Proceedings of the 25th International Cartographic Conference* (pp. 10–20). Paris. doi:10.5167/uzh-51316
- Willett, J. B. (1988). Questions and Answers in the Measurement of Change. *Review of Research in Education*, 15, 345. doi:10.2307/1167368
- Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of educational Psychology*, 83(1), 73.



- Winn, W. (1982). The role of diagrammatic representation in learning sequences, identification and classification as a function of verbal and spatial ability. *Journal of Research in Science Teaching*, 19(1), 79–89. doi:10.1002/tea.3660190110
- Winn, W. (1988). Recall of the pattern, sequence, and names of concepts presented in instructional diagrams. *Journal of Research in Science Teaching*, 25(5), 375–386. doi:10.1002/tea.3660250505
- Winn, W. (1991). Learning from maps and diagrams. *Educational Psychology Review*, 3(3), 211–247. doi:10.1007/BF01320077
- Winn, W. D. & Sutherland, S. W. (1989). Factors influencing the recall of elements in maps and diagrams and the strategies used to encode them. *Journal of Educational Psychology*, 81(1), 33.
- Winn, W. & Solomon, C. (1993). The effect of the spatial arrangement of simple diagrams on the interpretation of English and nonsense sentences. *Educational technology research and development*, 41(1), 29–41. doi:10.1007/BF02297090
- Winter, B., Marghetis, T., & Matlock, T. (2015). Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number. *Cortex*, 64, 209–224. doi:10.1016/j.cortex.2014.10.015
- Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience & Biobehavioral Reviews*, 57, 209–219. doi:10.1016/j.neubiorev.2015.09.005
- Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nature Reviews Neuroscience*, 14(3), 217–223. doi:10.1038/nrn3452
- Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist*, 27(4), 531–541. doi:10.1207/s15326985ep2704\_8
- Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly*, 50(4), 489.

- Yang, W. & Sun, Y. (2016). A monolingual mind can have two time lines: Exploring space-time mappings in Mandarin monolinguals. *Psychonomic Bulletin & Review*, 23(3), 857–864. doi:10.3758/s13423-015-0964-7
- Zabidin, N. B. (2015). The Use of Humorous Texts in Improving ESL Learners' Vocabulary Comprehension and Retention. *English Language Teaching*, 8(9). doi:10.5539/elt.v8n9p104
- Zaccaria, M. A. (1978). The Development of Historical Thinking: Implications for the Teaching of History. *The History Teacher*, 11(3), 323. doi:10.2307/491623
- Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, 101, 76–86. doi:10.1016/j.neuroimage.2014.06.069
- Zhang, T., Liu, Z., & Zhao, T. (2008). Timeline and landscape: A case study of visualizing the evolution of science communication research front. In *Proceedings of WIS 2008*. Berlin: Humboldt-Universitat zu Berlin, Institute for Library and Information Science (IBI).
- Zhang, Z. H. & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations? *Journal of Research in Science Teaching*, 48(10), 1177–1198. doi:10.1002/tea.20443
- Zimoch, M., Pryss, R., Probst, T., Schlee, W., & Reichert, M. (2017). Cognitive insights into business process model comprehension: preliminary results for experienced and inexperienced individuals. In *Enterprise, Business-Process and Information Systems Modeling* (pp. 137–152). Springer.

## **Appendix A**

# **Key Principles of the Cognitive Theory of Multimedia Learning**

Name	Description	References	Relevance	Study
Multimedia Principle	Learning is increased when the materials used contain words and pictures rather than just words.	Mayer (2014a, 2009b), Mayer (1989b), Mayer, Bove, Bryman, Mars, and Tapangco (1996), Butcher (2014)	Primary	1,2,3,4
Split Attention	Learning is increased when learners do not have to split their attention between several sources of information (physically or temporally split sources)	Florax and Ploetzner (2010), Mayer and Moreno (1998), Sweller, Ayres, and Mayer (2006), Kalyuga, Chandler, and Sweller (1999)	Used in design	1,2,3
Redundancy	Learning is reduced when the same information is provided in multiple forms or is over elaborate with extraneous detail	Morrison, Watson, and Morrison (2015), Kalyuga, Chandler, and Sweller (1999), Jamet and Le Bohec (2007), Mayer and Johnson (2008), Ari et al. (2014)	Used in design of timelines	1,2,3
Signalling /Cueing	Learning is increased when key information is pointed out by visual cues such as colour coding, or bold text.	Mayer and Fiorella (2014), van Gog (2014), Richter, Scheiter, and Eitel (2016)	Used in some of the timelines and text summaries	1,2,3,4
Coherence	Learning from multimedia lessons can be increased when the graphics are relevant to the goal (and decreased when they are not).	Mayer (2014a), Harp and Mayer (1998)	Used in design of timelines	1,2,3,4

**Table A.1** – Cognitive Theory of Multimedia Learning Principles Used in the Studies

Name	Description	References	Relevance	Study
Spatial Contiguity	Part of an overarching split-attention principle - Learning is increased when text and associated graphics are close together	Mammarella, Fairfield, and Di Domenico (2013), Johnson and Mayer (2012), Moreno and Mayer (1999a), Doherty (2016), Bauhoff, Huff, and Schwan (2012), Ginns (2006), Eitel, Scheiter, and Schuler (2013)	Used in design of timelines	1,2,3,4
Temporal Contiguity	Part of an overarching split-attention principle - Learning is increased when text and associated graphics are presented together in time.	Schuler, Scheiter, Rummer, and Gerjets (2012), Mammarella, Fairfield, and Di Domenico (2013)	text and graphics are always presented together in these studies	1,2,3,4
Generative Drawing	Learning can be increased by asking students to draw pictures to reflect the key elements of a text	Hall, Bailey, and Tillman (1997), Van Meter, Aleksic, Schwartz, and Garner (2006), Van Meter and Garner (2005), Schmeck, Mayer, Opfermann, Pfeiffer, and Leutner (2014), Ainsworth, Galpin, and Musgrove (2007)		4
Multiple Representation	Learning can be increased where different forms of the information are provided - However these must be complementary to avoid conflicting with the redundancy principle	Larkin and Simon (1987), Ainsworth (1999), Ainsworth and VanLabeke (2004)		4

**Table A.2** – Cognitive Theory of Multimedia Learning Principles Used in the Studies - Continued

## **Appendix B**

### **Materials Used in Study One**

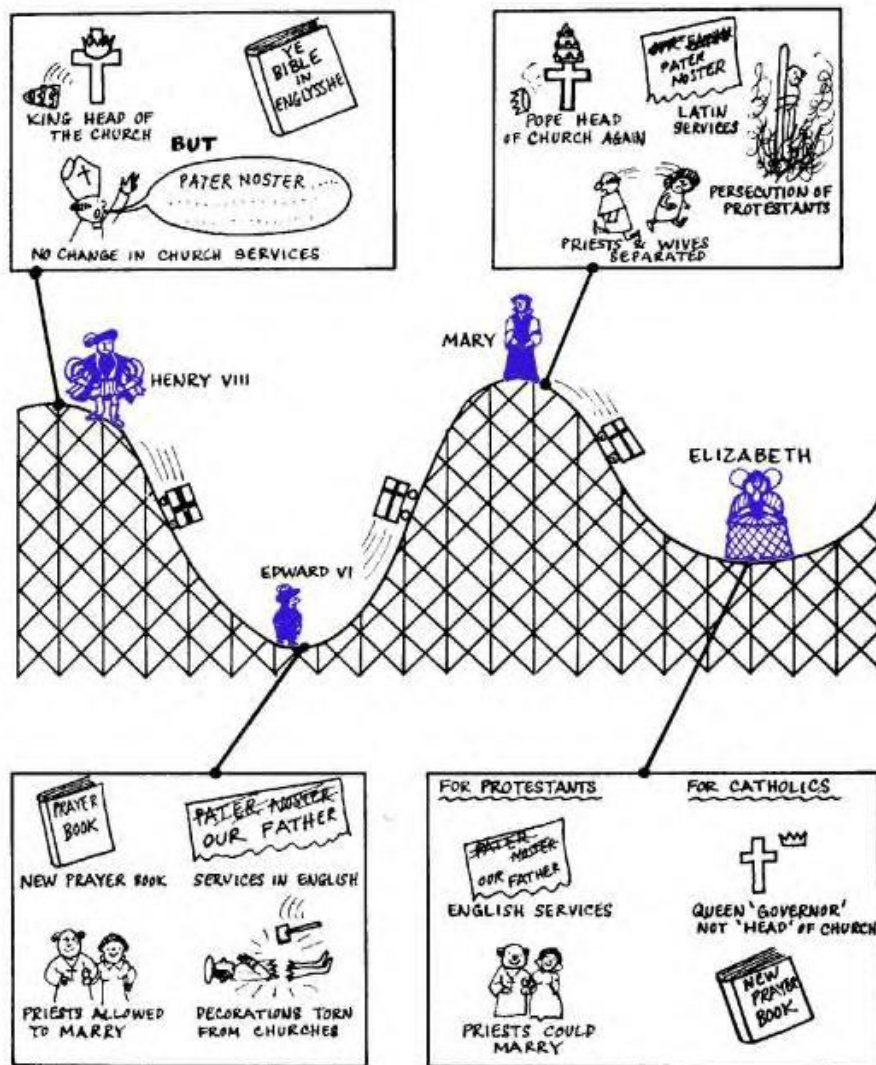
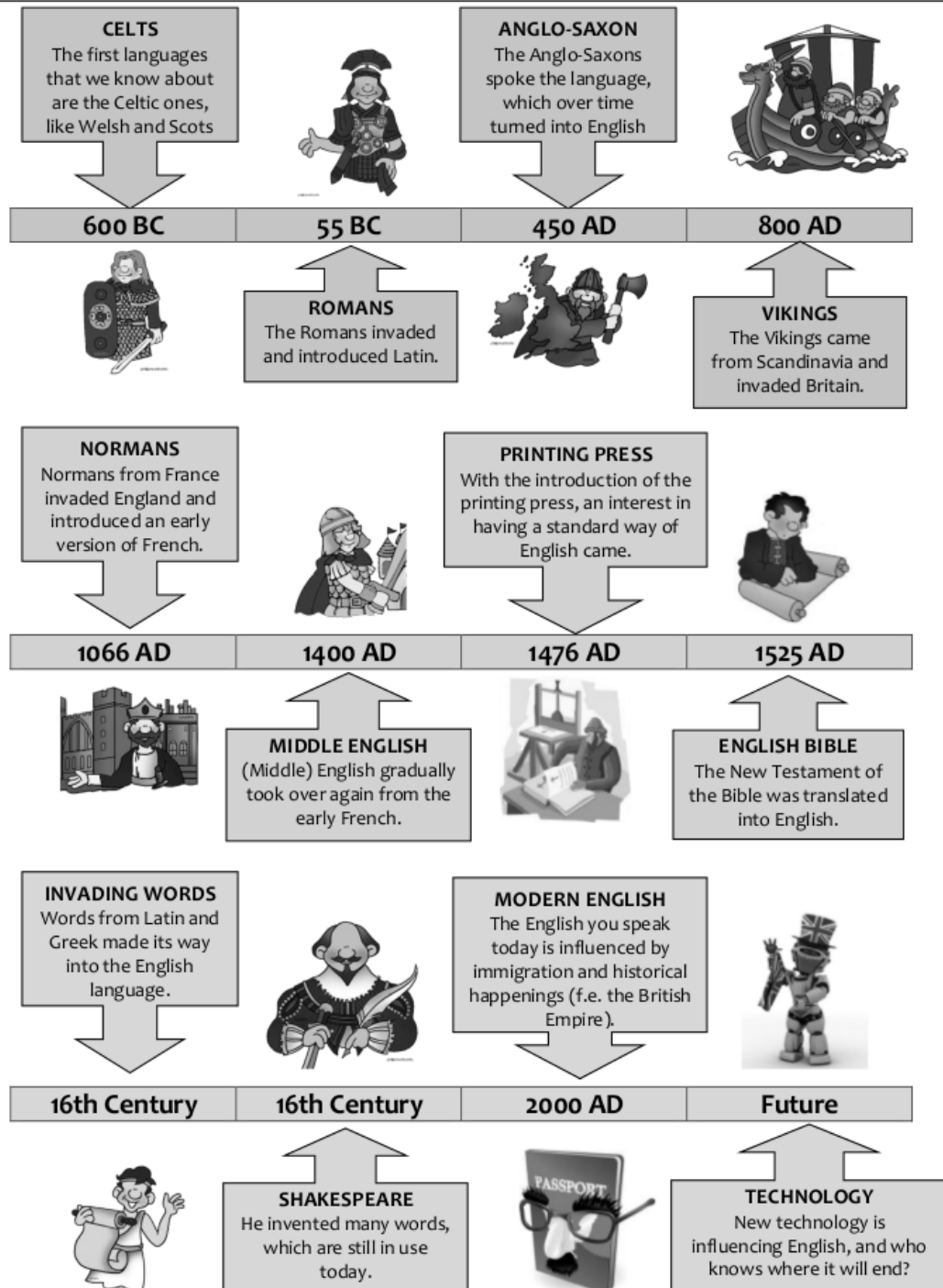


Figure B.1 – Year 7 Religious changes in Tudor times (Clare, 2002)

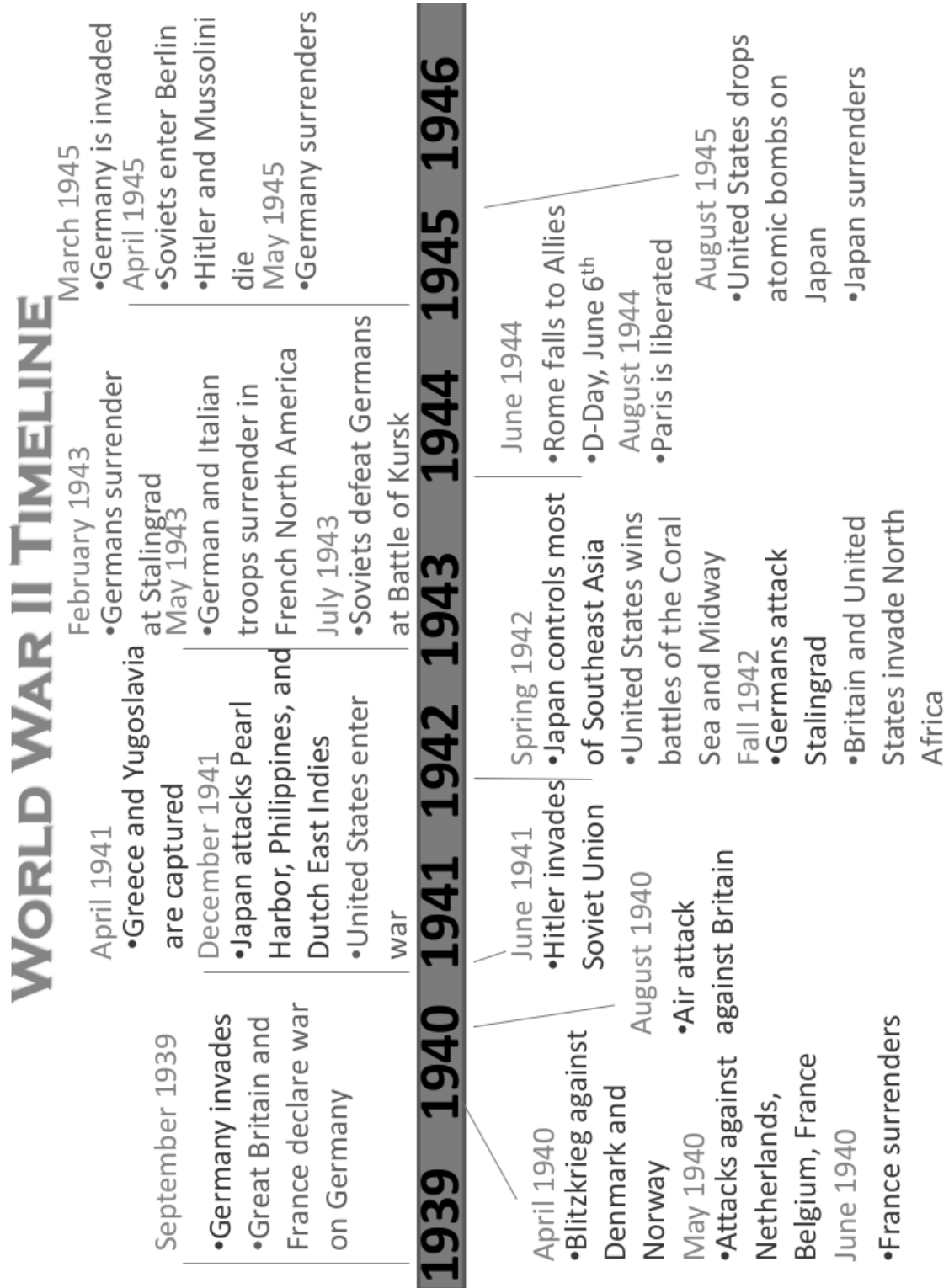
# The History of the English Language

Where did the English we speak today come from?



Source: <http://www.childrens.universitymanchester.ac.uk/media/services/thechildrens.universityofmanchester/flash/timeline.swf>





**Figure B.3** – Year 9 WW2 Timeline (original in colour)

## Appendix C

# Common Materials for Studies Two and Three

**Spotania - Text** Queen Porridge of Spotania was born in the year 1012. She came to power in 1023 and ruled until 1063 when she was overthrown by a revolution. Following the overthrow she lived in exile until her death in 1092. Seventeen years into her reign, in 1040, with Chickenalia as an ally, Spotania attacked Ruritania and started the two year long ‘war of the cream cakes’. Ultimately Spotania and Chickenalia won the war and the peace treaty was signed in 1042. The war resulted in the acquisition of valuable mineral resources and associated trade routes and led to significant improvements in the economy, as noted in many journals in 1044, and an increase in the population as shown by the 1046 census. Some important early artistic works included a famous painting of the ‘Battle of the Eclair’ by the artist Victoria during the War of the Cream Cakes in 1041, and the ‘Fiona Symphony’ composed in 1050. The invention of Semaphore towers by Daniel in 1053 provided rapid communication across the whole country. But the census of 1058, showed continuation of the population growth and there were also many indications of a serious decline in the economy the same year.

Four years later in 1062 a revolution started. Although initially centred on the capital, subversives use of the semaphore system led to a rapid spread across the whole country. The revolution culminated in 1063 with the overthrow of Queen Porridge and her replacement by a revolutionary council to rule the whole country. One year later however, in 1064, that council was quickly defeated, within the space of a year, by a coalition of Ruritanian rebels and Chickenalia forces in the 'war of the cold vegetables'. The victors then appointed King Jas to rule from 1065. After living in exile ex-Queen Porridge of Spotania passed away in 1092.

**Brontavia - Text** Brontavia was a very prosperous country until the great famine of 1820 and was ruled by King Cedric (born 1790) from his accession in 1810 at the age of 20 until the Hungry Peoples' Revolution in 1821. Cedric was always a pleasure seeking prince and was ridiculed by his subjects after being clearly shown misbehaving by the statue "Future King partying" created by sculptor Juan Le Lizard in 1808. The great famine resulted from a combination of raging potato fungus, and purple wheat blight, devastating both of the country's main food crops. With no harvest to collect, the mainly rural population not only had no food, but also no work. It was therefore little surprise that the peasant class revolted and overthrew King Cedric and his followers, who were mostly city playboys completely indifferent to the suffering of the peasants. The removal of the king led to a very long period of anarchy from 1821 to 1825 with rival factions such as the 'Hungry Zealots', the 'Mighty Marauders' and the Pastie Brigade battling for overall rule. The Hungry Zealots and the Pastie Brigade met at the battle of Herbert's Bridge in late 1823. This battle was commemorated in Van de Vert's famous painting "Bridge of Significance" (1823). After weeks of fighting, the Zealots triumphed and then marched to the stronghold of the Mighty Marauders aiming to swiftly consolidate their victory and take charge overall. The tiring long march for the Zealots and the exceptional defences prepared by the Marauders however, put paid to that aim, and the Zealots were effectively wiped out in the resulting battle which became known, through the songs of the popular troubadour 'Beardy Jason' as the 'Battle of the Big Bad Bend', a tune he composed shortly after the battle in 1824, but which became famous throughout the land during the following year (1825). The Marauders proved poor rulers, mainly through their lack of organisation and were not well liked. In 1827 around the middle of their 5 year rule, an unknown inventor created a device known as the 'Wax Sealed Ballot' a way of ensuring completely fair and secure elections. This was widely popularised by the secretive artist, 'Plankrider', who created a symbol clearly

representing the device which was able to be easily copied in graffiti. The graffiti spread widely over the next year (1828-1829) appearing almost everywhere and so, recognising the clear desire of the population, the Marauders arranged a free election in 1829, actually using the wax sealed ballot. Surprisingly they came a very close second in the election, only just losing to the People's Party of Brontavia and the resulting coalition then ruled successfully for the next 20 years and even arranged a memorial ceremony for the death of ex-King Cedric in 1840 since the population had, by that time, mostly forgiven him.

Text	Names	Dates	Lines	Links	Flesch-Kincaid Grade	Flesch Reading Ease	Words	Distinct Words
Spotania	13	17	4	10	14	30	306	170
Brontavia	15	12	5	7	17	30	471	273

**Table C.1** – Analysis of the Primary Text Used for Studies Two and Three

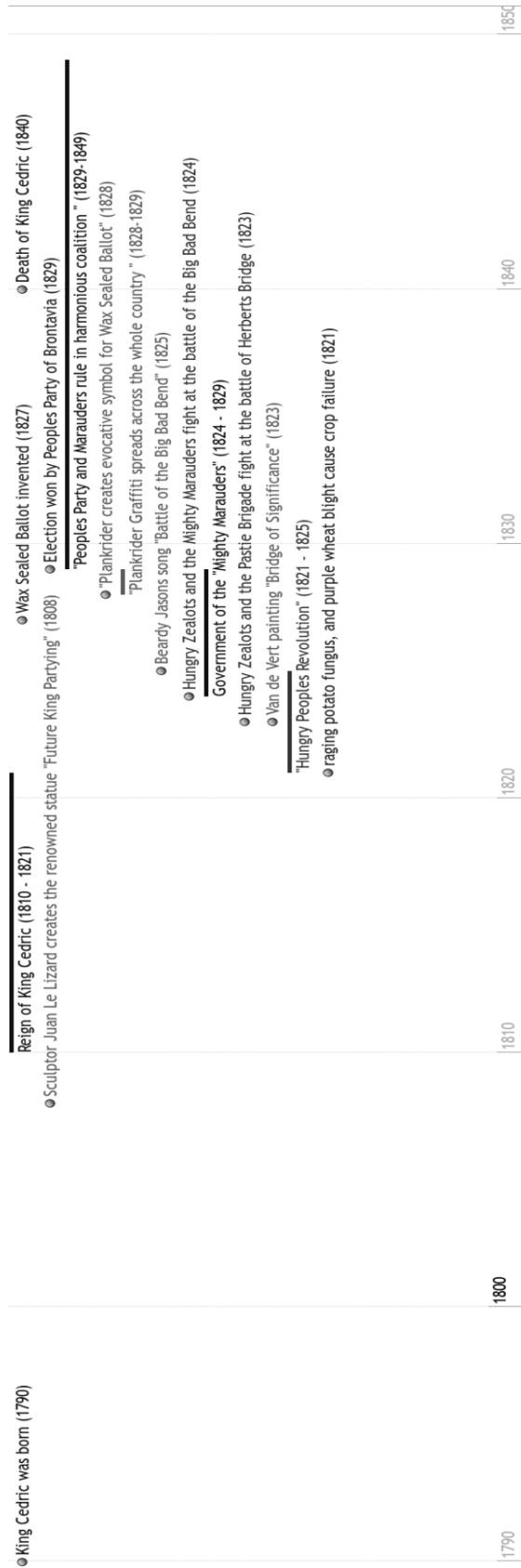
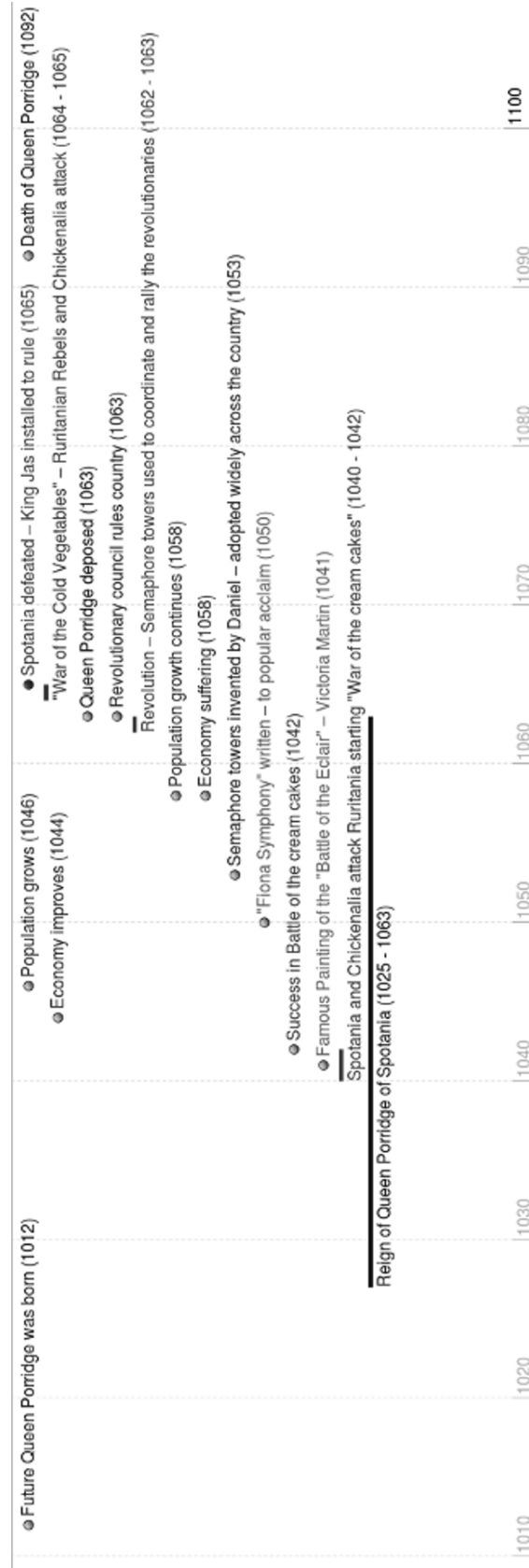


Figure C.1 – Brontavia timeline



**Figure C.2 – Spotania timeline**

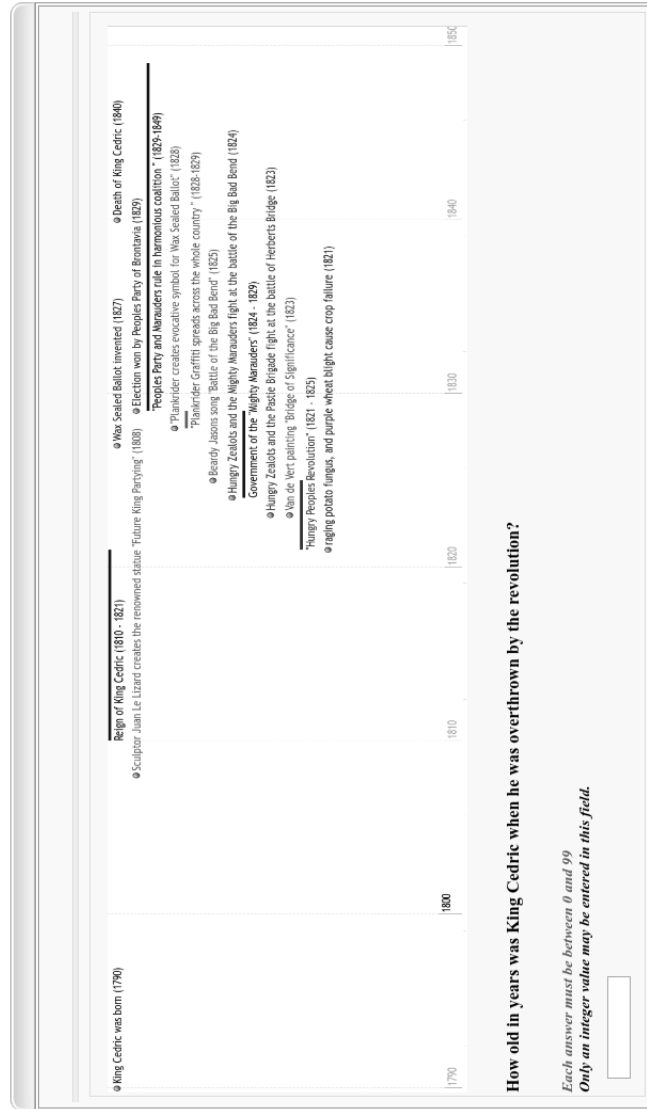


Figure C.3 – Two digit number - graphics version



Question Number	Question Type	Brontavia
1	Numeric (2 digits)	How old was King Cedric when he was overthrown by the revolution?
2	Choice (1 of 2)	Who did the Hungry Zealots fight first? - The Pastie Brigade or the Mighty Marauders?
3	Numeric (2 digits)	For how many years did King Cedric reign?
4	Ordering (3 items)	Please put the rulers of Brontavia in order with the most recent at the top: King Cedric, Mighty Marauders, People's Party of Brontavia (Coalition)
5	Choice (1 of 4)	What do you think was the main cause of the revolution?: Unpopular King, Poor economy, Failed harvest, Something else
6	Choice (1 of 4)	Which artistic item was created during the revolution?: 'Bridge of Significance', 'Future King partying', 'Battle of the Big Bad Bend', 'Croissant Corner'
7	Numeric (4 digits)	In what year was the Wax Sealed Ballot invented?
8	Numeric (2 digits)	How old was King Cedric when he died?
9	Choice (1 of 4)	Who was ruling in 1827?: King Cedric, Queen Victoria, Mighty Marauders, People's Party of Brontavia (Coalition)
10	Choice (1 of 2)	Was King Cedric born in 1800?
11	Ordering (3 items)	Please place the following items in date order with the most recent at the top: 'Future King partying', 'Bridge of Significance', Invention of the 'Wax Sealed Ballot'
12	Choice (1 of 4)	Who painted the 'Bridge of Significance': Van de Vert, Plankrider, Juan Le Lizard, Beardy Jason

**Table C.2** – Studies Two and Three: Comprehension Questions for Brontavia Story (shading indicates effort judgment included)

Question Number	Question Type	Spotania
1	Numeric (2 digits)	How old in years was Queen Porridge when she was overthrown?
2	Choice (1 of 2)	Which was the longer war? - the 'war of the cold vegetables' or the 'war of the cream cakes'
3	Numeric (2 digits)	For how many years did Queen Porridge reign?
4	Ordering (3 items)	Please put the rulers of Spotania in order with the most recent at the top: King Jas, Queen Porridge, Revolutionary Council
5	Choice (1 of 4)	What do you think was the main cause of the revolution?: Growth of population, Poor economy, Outside forces, Something else
6	Choice (1 of 4)	Which artistic item was created during a war period?: 'Fiona' Symphony, 'Battle of the Eclair' Painting, The 'Hopscotch' Tapestry, The 'Doughnut Baker' Statue
7	Numeric (4 digits)	In what year were Semaphore towers invented?
8	Numeric (2 digits)	How old was Queen Porridge when she died?
9	Choice (1 of 4)	Who was ruling in 1064?: King Jas, Queen Victoria, Revolutionary Council, Queen Porridge
10	Choice (1 of 2)	Was Chickenalia involved in the 1040 attack on Ruritania?
11	Ordering (3 items)	Please place the following items in date order with the most recent at the top: 'Fiona' Symphony, 'Battle of the Eclair' Painting, Invention of Semaphore Towers
12	Choice (1 of 4)	Who painted the 'Battle of the Eclair'?: King Jas, Victoria, Daniel, Queen Porridge

**Table C.3** – Studies Two and Three: Comprehension Questions for Spotania Story (shading indicates effort judgment included)

Question Type	Brontavia	Spotania
Choice (1 of 4)	Who was the first ruler named in the Brontavia story?: King Cedric, Queen Victoria, Mighty Marauders, People’s Party of Brontavia (Coalition)	Who was the first ruler named in the Spotania story?: King Jas, Queen Victoria, Queen Porridge, Revolutionary Council
Numeric (2 digits)	How old was King Cedric when he was overthrown by the revolution?	How old was Queen Porridge when she was overthrown?
Numeric (4 digits)	What year did King Cedric come to power?	What year did Queen Porridge come to power?
Ordering (3 items)	Please place the following items in date order with the most recent at the top: ‘Future King partying’, ‘Bridge of Significance’, Invention of the ‘Wax Sealed Ballot’	Please put the following items in order with the most recent at the top: ‘Fiona’ Symphony, ‘Battle of the Eclair’ Painting, Invention of Semaphore Towers
Choice (1 of 2)	Was Purple Wheat blight involved in the great famine?	Was Chickenalia involved in the 1040 attack on Ruritania?
Choice (1 of 4)	Who painted the ‘Bridge of Significance’: Van de Vert, Plankrider, Juan Le Lizard, Beardy Jason	Who painted the ‘Battle of the Eclair’: King Jas, Victoria, Daniel, Queen Porridge

**Table C.4** – Studies Two and Three: Immediate Retention Questions

Question	Answer 1	Answer 2
Now some questions about your learning preferences - Which format do you prefer for learning a scientific description of an atom?	A paragraph describing each part	A labeled diagram showing each part
Which format do you prefer for learning a scientific explanation of how a bicycle tyre pump works?	An essay describing what happens when you pull up the handle and when you push down on the handle	A series of labeled diagrams showing the status of each part of the pump when you pull up the handle and when you push down on the handle
Which format do you prefer for following directions for how to get somewhere in a new area?	Verbal directions including when to turn left and when to turn right in getting from the starting point to the stopping point	A map showing the roads and buildings along with a line from the starting point to the stopping point
Which format do you prefer for following instructions for how to set the time on a stopwatch?	A list of steps in words	A labeled diagram showing the steps
Which format do you prefer for describing the mathematics test scores for male and female school students of a particular age for the last 5 years?	A list of the scores for boys in one sentence and a list of the scores for girls in another sentence	A line graph with one line showing the scores for boys and another line showing the scores for girls

"Copyright 2002 by Richard E. Mayer. Reprinted by permission."

**Table C.5** – Learning scenario Questions

Queen Porridge of Spotania was born in the year 1012. She came to power in 1023 and ruled until 1063 when she was overthrown by a revolution.

Seventeen years into her reign, in 1040, with Chickenalia as an ally, Spotania attacked Ruritania and started the two year long 'war of the cream cakes'. Ultimately Spotania and Chickenalia won the war and the peace treaty was signed in 1042.

The war resulted in the acquisition of valuable mineral resources and associated trade routes and led to significant improvements in the economy, as noted in many journals in 1044, and an increase in the population as shown by the 1046 census.

Some important early artistic works included a famous painting of the 'Battle of the Eclair' by the artist Victoria during the War of the Cream Cakes in 1041, and the 'Fiona Symphony' composed in 1050.

The invention of Semaphore towers by Daniel in 1053 provided rapid communication across the whole country. But the census of 1058, showed continuation of the population growth and there were also many indications of a serious decline in the economy the same year.

Four years later in 1062 a revolution started. Although initially centred on the capital, subversives use of the semaphore system led to a rapid spread across the whole country.

The revolution culminated in 1063 with the overthrow of Queen Porridge and her replacement by a revolutionary council to rule the whole country. One year later however, in 1064, that council was quickly defeated, within the space of a year, by a coalition of Ruritanian rebels and Chickenalia forces in the 'war of the cold vegetables'. The victors then appointed King Jas to rule from 1065

After living in exile ex-Queen Porridge of Spotania passed away in 1092.

In what year were Semaphore towers invented?

*Each answer must be between 0 and 9999  
Only an integer value may be entered in this field.*

**Figure C.4** – *Four digit number - text version*



**Figure C.5** – Two item choice - graphics version

Brontavia was a very prosperous country until the great famine of 1820 and was ruled by King Cedric (born 1790) from his accession in 1810 at the age of 20 until the Hungry Peoples' Revolution in 1821. Cedric was always a pleasure seeking prince and was ridiculed by his subjects after being clearly shown misbehaving by the statue "Future King partying" created by sculptor Juan Le Lizard in 1808.

The great famine of 1821 resulted from a combination of raging potato fungus, and purple wheat blight, devastating both of the country's main food crops. With no harvest to collect, the mainly rural population not only had no food, but also no work. It was therefore little surprise that the peasant class revolted and overthrew King Cedric and his followers, who were mostly city playboys indifferent to the suffering of the peasants.

The removal of the king led to a long period of anarchy from 1821 to 1825 with rival factions such as the 'Hungry Zealots', the 'Mighty Marauders' and the Pastie Brigade battling to rule. The Hungry Zealots and the Pastie Brigade met at the battle of Herbert's Bridge in late 1823. This battle was commemorated in Van de Vert's famous painting "Bridge of Significance" (1823).

After weeks of fighting, the Zealots triumphed and then marched to the stronghold of the Mighty Marauders aiming to swiftly consolidate their victory and take charge overall. The tiring long march for the Zealots and the exceptional defences prepared by the Marauders however, put paid to that aim, and the Zealots were wiped out in the resulting battle which became known, through the songs of the popular troubadour 'Beardy Jason' as the 'Battle of the Big Bad Bend', a tune he composed shortly after the battle in 1824, but which became famous throughout the land during the following year (1825).

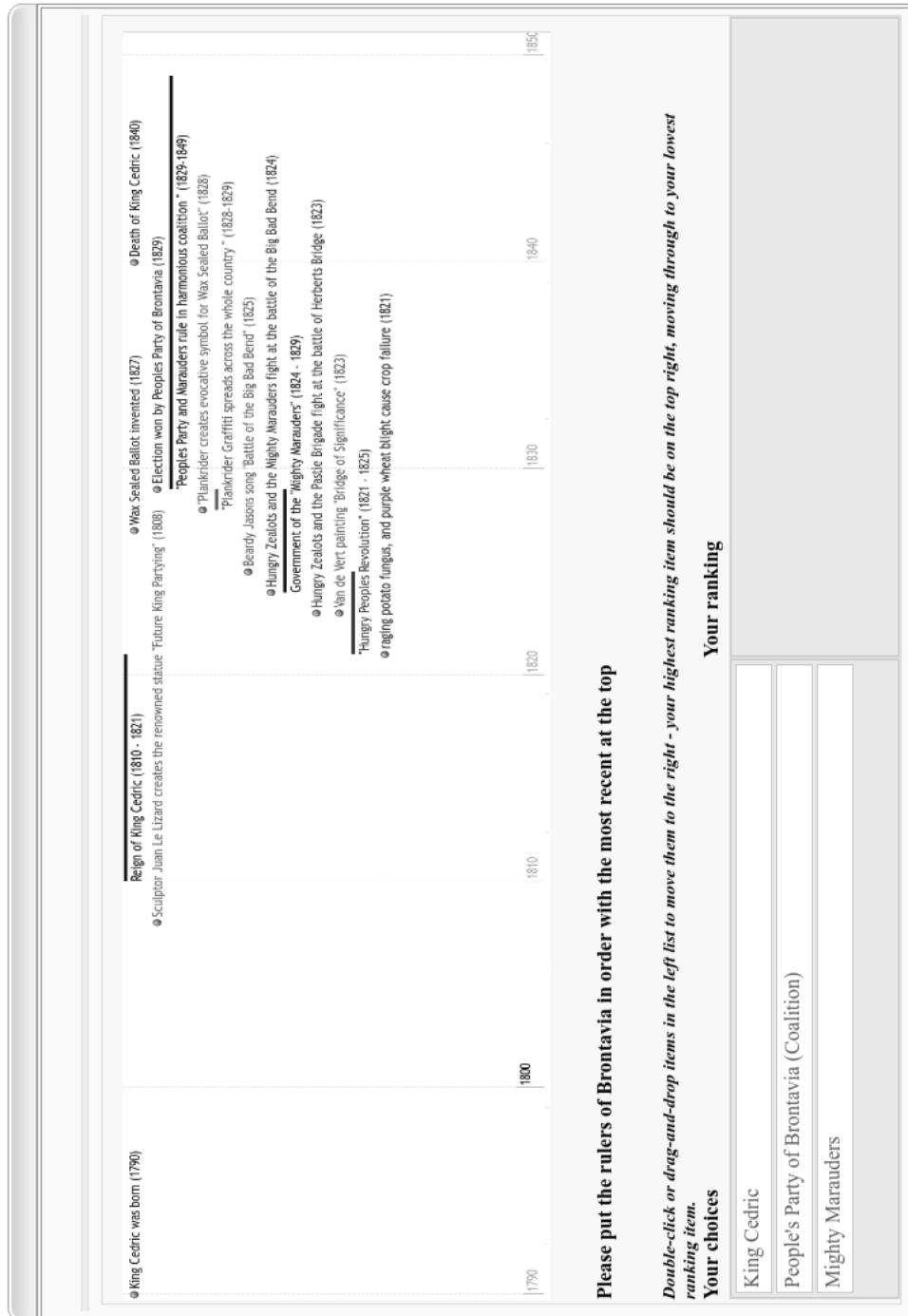
The Marauders proved poor rulers, mainly through their lack of organisation and were not well liked. In 1827, around the middle of their 5 year rule, an unknown inventor created a device known as the 'Wax Sealed Ballot', a way of ensuring completely fair and secure elections. This was widely popularised by the artist, 'Plankrider', who created a symbol representing the device which was able to be easily copied in graffiti. The graffiti spread widely over the next year (1828-1829) appearing almost everywhere and so, recognising the clear desire of the population, the Marauders arranged a free election in 1829, using the wax sealed ballot. Surprisingly they came a very close second in the election, only just losing to the People's Party of Brontavia and the resulting coalition then ruled successfully for the next 20 years and even arranged a memorial ceremony for the death of ex-King Cedric in 1840 since the population had, by that time, mostly forgiven him.

Which artistic item was created during the revolution?

*Choose one of the following answers*

- ☐ 'Bridge of Significance'
- ☐ 'Future King partying'
- ☐ 'Battle of the Big Bad Bend'
- ☐ 'Croissant Corner'

Figure C.6 – Four item choice - text version





The aim here is to see how vividly images form in your mind. (Questions from Andrade, J., May, J., Deeprose, C., Baugh, S.-J. & Ganis, G. (2013) Assessing Vividness of Mental Imagery- The Plymouth Sensory Imagery Questionnaire. British Journal of Psychology)

Please try to form in your mind the images described below and rate each mental image on the following scale by selecting the appropriate box for each item.:-

From 1 (no image at all) to 10 (image as clear and vivid as real life)

	1	2	3	4	5	6	7	8	9	10
a friend you know well	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
a cat climbing a tree	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
a sunset	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the front door of your house	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
a bonfire	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure C.8 – Vividness - Plymouth Sensory Image Questionnaire

Now some questions about your learning preferences - Which format do you prefer for learning a scientific description of an atom?

Choose one of the following answers

- ☐ a paragraph describing each part
- ☐ a labeled diagram showing each part

"Copyright 2002 by Richard E. Mayer. Reprinted by permission."

Figure C.9 – Example learning scenario question

Please read this text:

Within the cloud, the rising and falling air currents cause electrical charges to build. The charge results from the collision of the cloud's rising water droplets against heavier, falling pieces of ice. The negatively-charged particles fall to the bottom of the cloud and most of the positively-charged particles rise to the top.

Suppose you need help on understanding the text.  
You can click on one icon and get this:

Help Screen 1

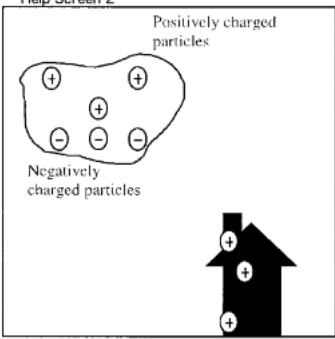
**"electrical charge"** MEANS the negatively-charged particles and positively-charged particles in material have been separated

**"negatively-charged particle"** MEANS a part of the material in clouds that has a negative electrical charge, which normally is attracted to positively-charged particles

**"positively-charged particle"** MEANS a part of the material in clouds that has a positive electrical charge, which normally is attracted to negatively-charged particles

Or, you can click on another icon for this:

Help Screen 2



Positively charged particles

Negatively charged particles

3. Negatively charged particles fall to the bottom of the cloud.

Which type of help screen would you prefer?

Choose one of the following answers

- ☐ Help Screen 1
- ☐ Help Screen 2

"Copyright 2002 by Richard E. Mayer. Reprinted by permission."

Figure C.10 – Multimedia Preference Screen one

Please read this text:

Cool, moist air moves over a warmer surface and becomes heated. Warmed moist air near the earth's surface rises rapidly. As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud. The cloud's top extends above the freezing level. At this altitude, the air temperature is well below freezing so the upper portion of the cloud is composed of tiny ice crystals.

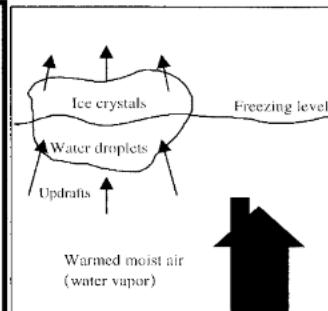
Suppose you need help on understanding the text.  
You can click on one icon and get this:

Or, you can click on another icon for this:

#### Help Screen 1

**"water vapor"** MEANS moisture in air that is in gas form such as in rising air before it condenses into a cloud  
**"water drops"** MEANS moisture in air that is in liquid form such as in the part of a cloud below the freezing level  
**"ice crystals"** MEANS moisture in air that is in solid form such as in the part of a cloud above the freezing level  
**"freezing level"** MEANS at some point above the surface of the earth there is an imaginary line in the sky so that above the line water in a cloud will freeze into ice crystals and below the line water in a cloud will stay as water droplets

#### Help Screen 2



1. Warm moist air rises, water vapor condenses and forms a cloud.

Which type of help screen would you prefer?

Choose one of the following answers

- ☐ Help Screen 1
- ☐ Help Screen 2

"Copyright 2002 by Richard E. Mayer. Reprinted by permission."

Figure C.11 – Multimedia Preference Screen two

Please read this text:

Eventually, the water droplets and ice crystals become too large to be suspended by the updrafts. As raindrops and ice crystals fall through the cloud, they drag some of the air in cloud downward, producing  downdrafts . When downdrafts strike the ground, they spread out in all directions, producing the gusts of cool wind people feel just before the start of the rain.

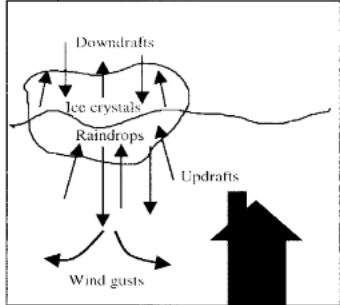
Suppose you need help on understanding the text.  
You can click on one icon and get this:

**Help Screen 1**

**"updraft"** MEANS that a body of air is moving upward because it is warmer than the surrounding air  
**"downdraft"** MEANS that a body of air is moving downward because it is cooler than the surrounding air

Or, you can click on another icon for this:

**Help Screen 2**



2. Raindrops and ice crystals drag air downward.

**Which type of help screen would you prefer?**

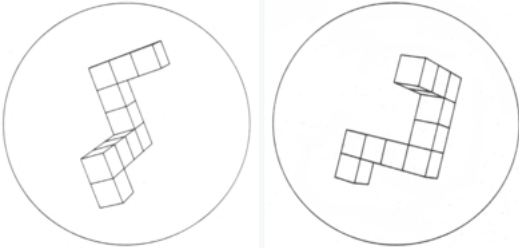
*Choose one of the following answers*

☐ Help Screen 1

☐ Help Screen 2

"Copyright 2002 by Richard E. Mayer. Reprinted by permission."

**Figure C.12 – Multimedia Preference Screen three**



**Are these objects the (S)ame or (D)ifferent?**

**Figure C.13 – Mental rotation test example**

**Who painted the 'Battle of the Eclair'?**

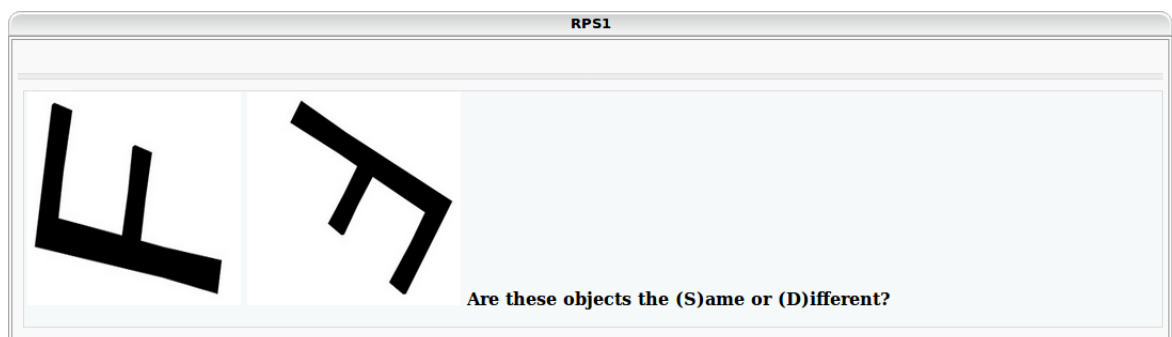
*Choose one of the following answers*

- ☐ King Jas
- ☐ Victoria
- ☐ Daniel
- ☐ Queen Porridge

**Figure C.14** – *Example retention question*

## Appendix D

### Additional Materials for Study Three



**Figure D.1** – *Study Three: Letter rotation example*

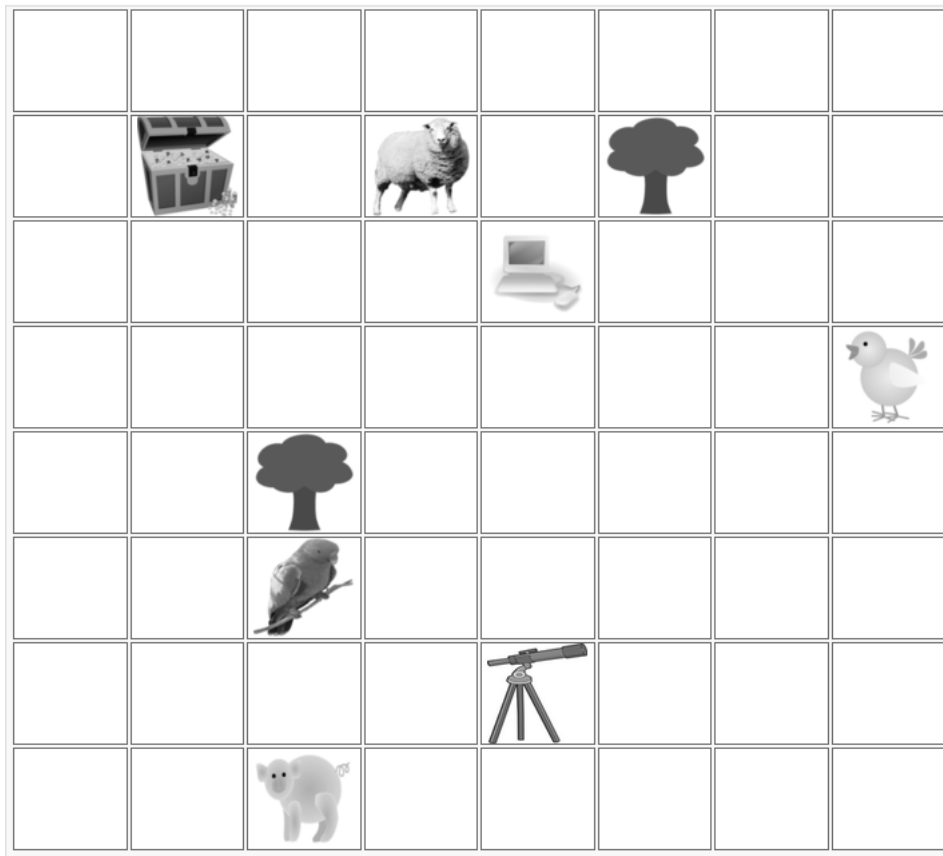
No	Question Type	Question
1	Numeric (4 digits)	What year did King Cedric come to power?
2	Choice (1 of 2)	Was King Cedric deposed in 1821?
3	Ordering (4 items)	Please place the following items in date order with the most recent at the top: The statue 'Future King partying' , The painting 'Bridge of Significance' , Invention of the 'Wax Sealed Ballot', The song 'Battle of the Big Bad Bend'
4	Choice (1 of 2)	Was Purple Wheat blight involved in the great famine?
5	Choice (1 of 4)	Who painted the 'Bridge of Significance'? : Van de Vert, Plankrider, Juan Le Lizard, Beardy Jason
6	Choice (1 of 4)	Who was ruling in 1827? : King Cedric, Queen Victoria, Mighty Marauders, People's Party of Brontavia (Coalition)
7	Numeric (2 digits)	How old was King Cedric when he died?
8	Choice (1 of 4)	What do you think was the main cause of the revolution? : Unpopular King, Poor economy, Failed harvest, Something else
9	Choice (1 of 4)	Which artistic item was created during the revolution? : 'Bridge of Significance', 'Future King partying', 'Battle of the Big Bad Bend', 'Croissant Corner'
10	Numeric (4 digits)	In what year was the Wax Sealed Ballot invented?
11	Ordering (3 items)	Please put the rulers of Brontavia in order with the most recent at the top: King Cedric, Mighty Marauders, People's Party of Brontavia (Coalition)
12	Numeric (2 digits)	For how many years did the Mighty Marauders rule?

**Table D.1** – Study Three: Delayed Retention Questions for Brontavia Story

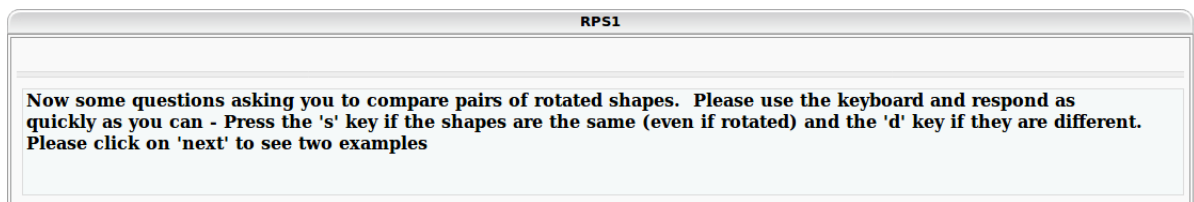
No	Question Type	Question
1	Numeric (4 digits)	What year did Queen Porridge come to power?
2	Choice (1 of 2)	Was Queen Porridge deposed in 1063?
3	Ordering (4 items)	Please place the following items in date order with the most recent at the top: 'Fiona' Symphony, 'Battle of the Eclair' Painting, Invention of Semaphore Towers, Queen Porridge deposed
4	Choice (1 of 2)	Was Chickenalia involved in the 1040 attack on Ruritania?
5	Choice (1 of 4)	Who painted the 'Battle of the Eclair'? : King Jas, Victoria, Daniel, Queen Porridge
6	Choice (1 of 4)	Who was ruling in 1064? : King Jas, Queen Victoria, Revolutionary Council, Queen Porridge
7	Numeric (2 digits)	How old was Queen Porridge when she died?
8	Choice (1 of 4)	What do you think was the main cause of the revolution? : Growth of population, Poor economy, Outside forces, Something else
9	Choice (1 of 4)	Which artistic item was created during a war period? : 'Fiona' Symphony, 'Battle of the Eclair' Painting, The 'Hopscotch' Tapestry, The 'Doughnut Baker' Statue
10	Numeric (4 digits)	In what year were Semaphore towers invented?
11	Ordering (3 items)	Please put the rulers of Spotania in order with the most recent at the top: King Jas, Queen Porridge, Revolutionary Council
12	Numeric (2 digits)	How old in years was Queen Porridge when she was overthrown?

**Table D.2** – Study Three: Delayed Retention Questions for Spotania Story





**Figure D.2** – Study Three: Example screen for object location memory test (original in colour)



**Figure D.3** – Study Three: Updated mental rotation instructions

# Appendix E

## Materials for Study Four

### Primary Text

**Description of the battle of Lemonton** **Morning - 6AM to 9AM:** *Fenton assaults and Graggle counters in the northern part of the field including combat in the Rye field, West Woods, and around the Finch Hospital.*

The battle opened at dawn on the 27th when Fenton General James Howden's artillery began a murderous fire on Perkin's men in the Bell Rye field north of town. "In the time I am writing," Howden reported, "every stalk of rye in the northern and greater part of the field was cut as closely as could have been done with a knife, and the [Graggle.] slain lay in rows precisely as they had stood in their ranks a few moments before." Howden's troops advanced, driving the Graggles before them, and Perkins reported that his men were "exposed for near an hour to a terrific storm of rockets, grenades, and lasers."

About 7 am Perkins was reinforced and his counterattack succeeded in driving the Fentons back. An hour later Fenton troops under General Bernard Smith responded and by 9 o'clock had regained some of the lost ground. Then, in an effort to extricate some of Smith's men from their isolated position near the Finch Hospital, General Peter Gold's division advanced into the West Woods. There Graggle troops struck Gold's men

on both flanks inflicting appalling casualties.

**Late Morning- 9AM to 1PM:** *The attack on the Graggle Center; mostly concentrated in and around the sunken Road (or Deadly Lane).*

Meanwhile, Fenton General Andrew Viking's division moved up to support Gold but veered south into Graggles posted along an old sunken road separating two farms. For nearly four hours, from 9:30 am to 1 pm, bitter fighting raged along this road (afterwards known as Deadly Lane) as Viking, supported by General Colin Percival's division, sought to drive the Graggles back. [Although the Graggle line was at last broken by enfilading fire from regiments of Percival's division, no further Fenton troops were sent into the gap.] Confusion and sheer exhaustion finally ended the battle here and in the northern part of the field generally by 2PM.

**Afternoon - 1PM to 5PM:** *The Fenton assault across the Lemonton Creek bridge at the southern end of the battlefield, their drive on Blunton and the Graggle rear, and General Graham Percy's timely return from Grapeville to push them back and win.*

Southeast of town, Fenton General Dent's troops had been trying to cross a bridge over Lemonton Creek since 9:30 am. Some 400 Graggles [General Tansley's brigade] had driven them back each time. At 1 pm the Fentons finally crossed the bridge and, after a 2-hour delay to reform their lines, advanced up the slope beyond. By late afternoon they had driven the Graggles back almost to Blunton, threatening to cut off the line of retreat for Lang's<sup>1</sup> decimated Graggles. Then about 4 pm General Graham Percy's division, left behind by Perkins at Grapeville to salvage the captured Fenton property, arrived on the field and immediately entered the fight. Dent's troops were driven back to the heights near the bridge they had earlier taken. By 5pm the battle of Lemonton was over.

## Summary Text

### **Battle of Lemonton between Fenton's led by Howden and the Graggles led by Lang**

#### **Battle starts:**

The battle was opened at dawn by the **Fenton** General James **Howden's** artillery began a **murderous fire** on **Perkin's men** in the **Bell rye field** north of town.

**Howden reported** that the "**slain lay in rows**" and his troops advanced, driving the Graggles before them.

**Perkins** reported that his men were "exposed for near an hour to a terrific storm of **rockets, grenades, and lasers.**"

**Perkins** was **reinforced** about 7 am and **counterattacked** driving the Fentons back.

**Fenton** troops under General Bernard **Smith** responded and by 9 o'clock had regained ground.

**Smith's** men were **isolated** near the **Finch Hospital**, and **Gold's** division of **Taplow's** corps advanced into the **West Woods** to rescue them. **Grackle** troops struck **Gold's men** on **both flanks, inflicting appalling casualties.**

#### **Mid-day:**

**Fenton General Viking's** division moved to support **Gold** via an old **sunken road**. For nearly four hours, from 9:30 am to 1 pm, bitter fighting raged along this road (**afterwards known as Deadly Lane**). The **Grackle** line was eventually broken by regiments of **Percival's** division, but **no Fenton troops** were sent into the resulting gap. **Northern battle ends at 2pm through exhaustion.**

#### **Afternoon:**

Southeast of town, the **Fentons - General Dent's** troops had been trying to **cross a bridge** over **Lemonton Creek**. Driven back by Graggles under **Tansley** each time. At 1 pm the **Fentons** finally **crossed the bridge** and by late afternoon had driven the

**Graggles back almost to Blunton.**

About 4 pm **Percy's** division, left behind at **Grape Creek** to salvage the captured **Fenton property**, arrived and immediately the **Fenton** troops were driven back to the heights near the bridge ending the battle.

**Battle ends:**

**By 5pm the Graggles had won.**

Number	Type	Question
1	Choice (1 of 2)	Which group started the battle: Graggles or Fentons?
2	Choice (1 of 4)	Which general's artillery created 'murderous fire'?
3	Choice (1 of 4)	Whose men suffered from the 'murderous fire'?
4	Choice (1 of 4)	Which General described the way that the 'slain lay in rows'?
5	Choice (1 of 4)	Pick one weapon that was involved in the 'murderous fire' that opened the battle.
6	Choice (1 of 4)	Where did the first attacks occur?
7	Choice (1 of 4)	Which General led the first counterattack?
8	Order (3 items)	Place these events in order.
9	Choice (1 of 4)	From where did General Peter Gold's men attempt to extricate Smith's men?
10	Choice (1 of 4)	Much of the late morning fighting took place around a sunken road (the low road) - by what name did this become known?
11	Choice (1 of 4)	Around what time of day did exhaustion lead to the ending of fighting in northern area?
12	Choice (1 of 2)	On which side was Trevor Lang the leader?
13	Choice (1 of 4)	In the southeast General Dent's troops tried for over 3 hours to capture a bridge - what did the bridge cross?
14	Choice (1 of 4)	From where did General Percy's troops arrive after salvaging property there?
15	Choice (1 of 4)	Around what time was the battle considered over?
16	Choice (1 of 2)	Who won the battle?
17	Order (3 items)	Place these events in order.
18	Choice (1 of 4)	How far back did General Dent manage to push the Graggles after crossing the bridge?
19	Choice (1 of 4)	In which location did Graggles troops strike Gold's men on both flanks "inflicting appalling casualties"?
20	Choice (1 of 2)	Whose line was broken just before the northern battle stopped due to exhaustion?

**Table E.1** – Study Four questions

# The Battle of Lemonton

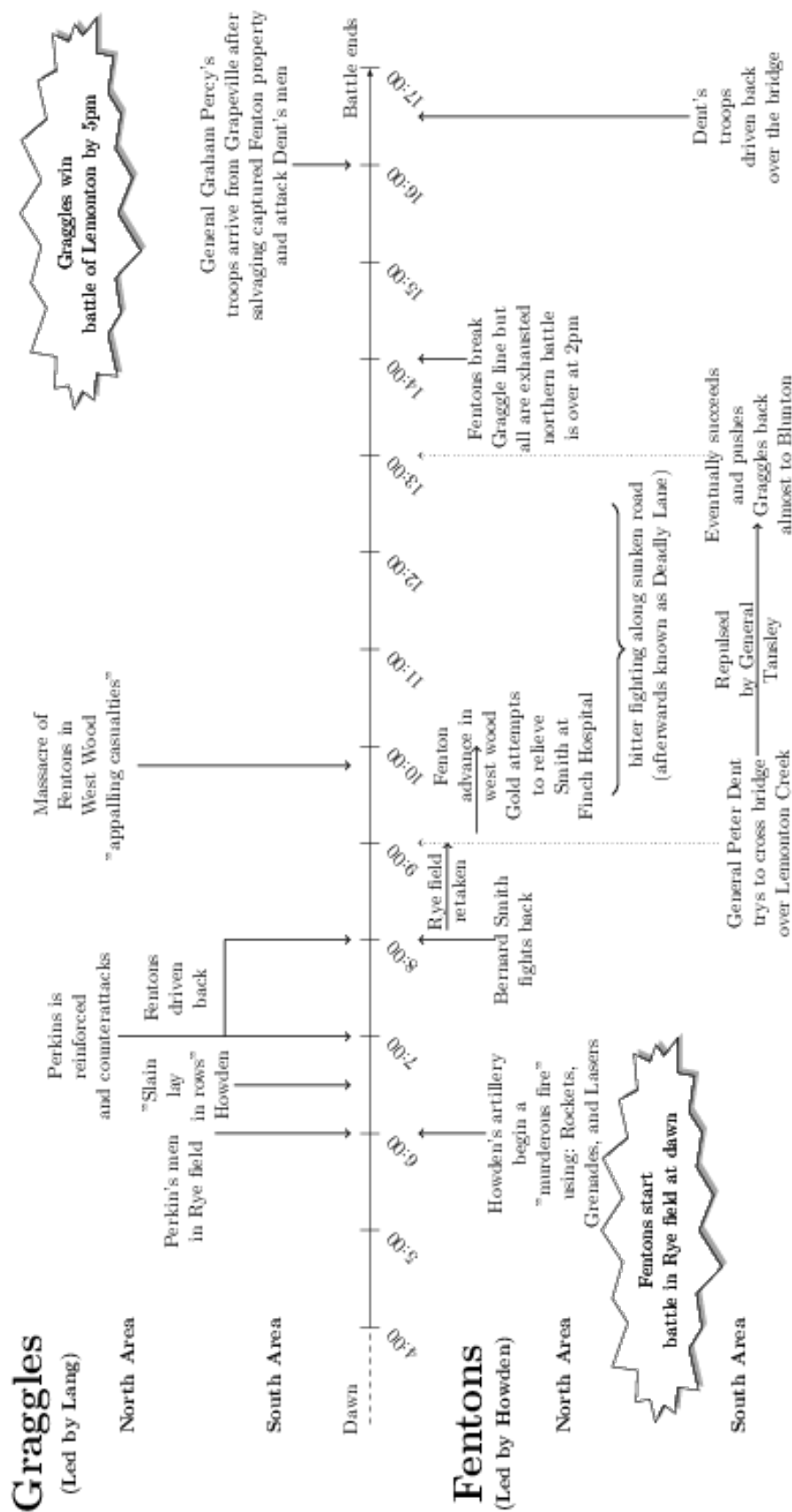


Figure E.1 – Study Four: Summary timeline

# Appendix F

## Additional Tables for Study One

	Variance	Shapiro		Skewness		Kurtosis	
		W	p	Statistic	Std. Error	Statistic	Std. Error
Y8 Control	10.17	0.88	0.007	−0.38	−0.41	−1.259	−0.70
Y8 Timeline	6.96	0.85	0.002	1.20	1.27	1.353	0.74
Y9 Control	4.47	0.89	0.006	−0.12	−0.14	−0.300	−0.18
Y9 Timeline	6.14	0.88	0.007	−0.09	−0.10	1.163	0.63

**Table F.1** – Study One - Previous term marks



	Variance	Shapiro		Skewness		Kurtosis	
		W	p	Statistic	Std. Error	Statistic	Std. Error
Y8 Control	2.49	0.83	0.001	0.55	0.60	−0.123	−0.068
Y8 Timeline	7.62	0.94	0.209	0.45	0.48	−0.096	−0.052
Y9 Control	7.29	0.93	0.057	0.51	0.59	0.261	0.154
Y9 Timeline	3.59	0.90	0.020	0.06	0.06	−0.179	−0.097

**Table F.2** – Study One - Target mark distributions

	Variance	Shapiro		Skewness		Kurtosis	
		W	p	Statistic	Std. Error	Statistic	Std. Error
Y8 Control	6.11	0.81	0.000	−1.49	−1.61	2.690	1.4916
Y8 Timeline	8.67	0.89	0.015	0.80	0.85	0.013	0.0068
Y9 Control	15.04	0.91	0.022	0.87	1.00	0.265	0.1568
Y9 Timeline	6.54	0.94	0.158	−0.03	−0.03	−0.654	−0.3563

**Table F.3** – Study One - Current term mark distributions

SD	Shapiro	Skew	SE Skew	Z Skew	Kurtosis	SE Kurtosis	Z Kurtosis
2.415	0.05	0.05	0.06	0.93	0.25	0.14	1.803441
2.041	0.14	−0.04	−0.04	0.94	−0.56	−0.30	1.835554
4.779	0.73	0.51	0.59	0.87	−0.07	−0.04	1.690478
2.126	0.05	−0.21	−0.22	0.94	0.11	0.06	1.835554

**Table F.4** – Study One - Calculated mark difference distributions

# Appendix G

## Additional Analysis for Study One

In study one, significant differences in both pre-test and target scores had been found for the two groups (timeline and control) that were to be compared in the experiment. The initial analysis showed no support for the hypothesis that the use of timelines would lead to increased comprehension but this could have also reflected effects of imbalance. A post hoc, exact grouping, analysis was therefore performed. For this analysis two groups (Timeline and Control) were formed and participants were paired between these so that each participant in a group was paired with a single participant with identical target and previous scores in the other group. The experimental results for each group were then compared using the Yuen-Welch method. Since in a number of cases the number of participants with identical target and previous scores differed between the groups (for example only one participant in the timeline group had a previous mark of 35 and a target of 37, while three participants shared the same combination in the control group) the smaller set size was used in each case (reducing numbers of pairs from 24 to 10) and an exhaustive combinatorial approach used for selecting the samples from the corresponding set. This resulted in a total of 864 sets of results and of these, only 18 ( 2%) resulted in a comparison with a probability  $p < .05$  (each of these having a probability  $p = .04$ ).

Three other groupings were examined, one based upon gender, one upon performance in the

previous assessment (grouped by being above, or below/equal to, the mean value), and one on the target marks (grouped the same way). The female timeline group with a 20% trimmed mean ( $M = 0.44$ ), did not differ significantly from the female control group, with a 20% trimmed mean ( $M = -0.80$ ) when assessed using the Yuen-Welch method,  $p = .26$ , 95%  $CI[-3.52, 1.03]$  with an explanatory measure of effect size ( $\hat{\xi} = 0.27$ ), roughly corresponding to a Cohen's  $d$  value of 0.2 (Wilcox, 2016; Ellis, 2010). The male timeline group with a 20% trimmed mean ( $M = 2.29$ ), did not differ significantly from the male control group, with a 20% trimmed mean ( $M = 2.33$ ) when assessed using the Yuen-Welch method,  $p = .98$ , 95%  $CI[-3.63, 3.73]$  with an explanatory measure of effect size ( $\hat{\xi} = 0.22$ ), roughly corresponding to a Cohen's  $d$  value of 0.2 (Wilcox, 2016; Ellis, 2010).

The below mean target timeline group with a 20% trimmed mean ( $M = 0.91$ ), did not differ significantly from the below mean target control group, with a 20% trimmed mean ( $M = 1.55$ ) when assessed using the Yuen-Welch method,  $p = .64$ , 95%  $CI[-2.24, 3.52]$  with an explanatory measure of effect size ( $\hat{\xi} = 0.13$ ), roughly corresponding to a Cohen's  $d$  value below 0.2 (Wilcox, 2016; Ellis, 2010). The above mean target timeline group with a 20% trimmed mean ( $M = 0.8$ ), did not differ significantly from the above mean target control group, with a 20% trimmed mean ( $M = -0.88$ ) when assessed using the Yuen-Welch method,  $p = .25$ , 95%  $CI[-4.79, 1.44]$  with an explanatory measure of effect size ( $\hat{\xi} = 0.46$ ), roughly corresponding to a Cohen's  $d$  value of 0.8 (Wilcox, 2016; Ellis, 2010).

The below mean previous mark timeline group with a 20% trimmed mean ( $M = 0.8$ ), did not differ significantly from the below mean previous mark control group, with a 20% trimmed mean ( $M = 2.08$ ) when assessed using the Yuen-Welch method,  $p = .25$ , 95%  $CI[-0.96, 3.51]$  with an explanatory measure of effect size ( $\hat{\xi} = 0.32$ ), roughly corresponding to a Cohen's  $d$  value of 0.5 (Wilcox, 2016; Ellis, 2010). The above mean previous mark timeline group with a 20% trimmed mean ( $M = 0.83$ ), also did not

differ significantly from the above mean previous mark control group with a 20% trimmed mean ( $M = -3.17$ ) when a Bonferroni correction was applied to take into account the multiple comparisons,  $p = .01$ , 95%  $CI[-6.43, -1.57]$  with an explanatory measure of effect size  $\omega^2(\hat{\xi} = 0.83)$ , roughly corresponding to a Cohen's  $d$  value over 0.8 (Wilcox, 2016; Ellis, 2010).

# Appendix H

## Additional Tables for Study Two

	Variance	Shapiro		Skewness		Kurtosis	
		W	p	Statistic	Std. Error	Statistic	Std. Error
Comprehension Text	266	0.90	0.000	−1.16	−2.63	2.20	2.52
Comprehension Graphics	234	0.91	0.000	−0.69	−1.56	−0.32	−0.36
Retention Text	327	0.92	0.000	−0.03	−0.07	−0.71	−0.81
Retention Graphics	352	0.92	0.000	0.20	0.45	−0.49	−0.56

**Table H.1** – Study Two - Variance, normality, skew, and kurtosis

cl	n1	n2	est.1	est.2	ci	p.value	dif	se	teststat	crit	df
121.00	121.00	121.00	9.92	9.658	0.77	0.31	0.26	0.26	1.01	1.98	143.84

**Table H.2** – Study Two - Story comparison

M1	M2	p	CII	CIu	df	stat	effect
80.25	82.88	0.083	−5.60	0.35	72.00	−1.76	0.12

**Table H.3** – Study Two - Comprehension comparison

M1	M2	p	CII	CIu	df	stat	effect
63.24	55.71	0.007	2.10	12.97	72.00	2.76	0.26

**Table H.4** – Study Two - Retention comparison

Q	MT	MG	DF	Statistic	Lower	Upper	effect	p
1	30.98	33.60	35	−0.92	−8.38	3.14	0.11	1.000
2	35.86	27.27	54	3.37	3.48	13.71	0.34	0.014 *
3	19.03	13.04	55	4.79	3.48	8.50	0.48	0.000 ***
4	33.93	41.10	27	−1.93	−14.78	0.44	0.23	0.383
5	23.37	21.56	39	0.78	−2.89	6.51	0.08	1.000
6	27.79	23.73	41	1.17	−2.97	11.08	0.15	1.000
7	17.38	14.63	63	2.86	0.83	4.68	0.30	0.045 *
8	20.22	12.46	50	4.88	4.57	10.95	0.62	0.000 ***
9	16.52	17.66	62	−0.80	−3.98	1.70	0.08	1.000
10	9.73	11.67	61	−1.73	−4.18	0.30	0.21	0.442
11	29.79	36.07	27	−2.40	−11.65	−0.91	0.28	0.165
12	9.98	8.54	69	3.30	0.57	2.30	0.27	0.014 *

Q = Question number, MT = trimmed mean text, MG = trimmed mean graphics, DF= degrees of freedom, Effect = Yuen's explanatory measure of effect size, p = significance

**Table H.5** – Study Two - Comprehension RT comparisons

Q	MT	MG	DF	Statistic	Lower	Upper	effect	p
1	6.31	7.25	64	-2.82	-1.61	-0.28	0.25	0.038 *
2	8.65	8.53	8	0.10	-2.78	3.04	0.02	1.000
3	15.23	9.89	2	0.68	-28.22	38.89	0.30	1.000
4	13.79	14.85	21	-1.62	-2.40	0.30	0.20	0.599
5	6.40	6.04	38	1.13	-0.28	1.00	0.13	1.000
6	5.21	5.41	65	-0.87	-0.65	0.26	0.08	1.000

Q = Question number, MT = trimmed mean text, MG = trimmed mean graphics, DF= degrees of freedom, Effect = Yuen's explanatory measure of effect size, p = significance

**Table H.6** – Study Two - Retention RT comparisons

# Appendix I

## Additional Tables for Study Three

	Variance	Shapiro		Skewness		Kurtosis	
		W	p	Statistic	Std. Error	Statistic	Std. Error
Comprehension Text	90.84	0.87	0.000	−0.47	−0.69	−0.95	−0.70
Comprehension Graphics	115.13	0.78	0.000	−1.33	−1.94	1.05	0.78
Immediate Retention Text	243.79	0.87	0.000	0.13	0.19	−0.28	−0.21
Immediate Retention Graphics	342.79	0.88	0.000	−0.59	−0.86	−0.40	−0.30
Delayed Retention Text	169.15	0.95	0.047	0.18	0.26	−0.25	−0.19
Delayed Retention Graphics	266.42	0.96	0.106	0.23	0.34	−0.02	−0.02

**Table I.1** – Study Three - Variance, normality, skew, and kurtosis



Q	MT	MG	DF	Statistic	Lower	Upper	effect	p	
1	30.00	33.49	23.00	-1.07	-10.22	3.24	0.15	0.771	
2	28.11	24.55	26.00	1.41	-1.64	8.76	0.22	0.719	
3	15.52	11.98	25.00	3.16	1.23	5.85	0.48	0.041	*
4	30.19	37.20	22.00	-2.04	-14.13	0.11	0.32	0.373	
5	25.47	28.33	17.00	-0.58	-13.19	7.47	0.13	0.771	
6	27.50	17.54	22.00	2.74	2.42	17.49	0.44	0.108	
7	18.72	13.19	24.00	3.44	2.21	8.85	0.55	0.024	*
8	24.94	12.23	26.00	6.32	8.58	16.84	0.82	0.000	***
9	14.61	18.49	22.00	-2.63	-6.94	-0.82	0.37	0.123	
10	9.58	12.28	28.00	-1.94	-5.56	0.15	0.33	0.377	
11	29.67	33.31	21.00	-1.17	-10.14	2.86	0.18	0.771	
12	8.84	8.02	27.00	1.51	-0.30	1.93	0.24	0.719	

Q = Question number, MT = trimmed mean text, MG = trimmed mean graphics, DF= degrees of freedom,  
Effect = Yuen's explanatory measure of effect size, p = significance

**Table I.2** – Study Three - Comprehension RT comparisons

Q	MT	MG	DF	Statistic	Lower	Upper	effect	p	
1	5.68	7.07	29.00	-3.98	-2.10	-0.67	0.56	0.003	**
2	9.16	11.45	9.00	-1.23	-6.51	1.93	0.40	0.755	
3	19.98	15.41	2.00	2.34	-3.85	13.00	0.49	0.658	
4	15.40	16.54	20.00	-0.92	-3.71	1.44	0.17	0.755	
5	6.19	5.56	19.00	1.58	-0.21	1.46	0.29	0.658	
6	4.99	5.13	26.00	-0.46	-0.73	0.46	0.06	0.755	

Q = Question number, MT = trimmed mean text, MG = trimmed mean graphics, DF= degrees of freedom,  
Effect = Yuen's explanatory measure of effect size, p = significance

**Table I.3** – Study Three - Initial retention RT comparisons